BMJ open

Quality improvement needed in quality improvement randomized trials: Systematic review of interventions to improve care in diabetes

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2013-002727
Article Type:	Research
Date Submitted by the Author:	14-Feb-2013
Complete List of Authors:	Ivers, Noah; Women's College Hospital, Family and Community Medicine Tricco, Andrea; Li Ka Shing Knowledge Institute of St Michael's Hospital Taljaard, Monica; Ottawa Health Research Institute, Halperin, Ilana; Sunnybrook Health Sciences Centre, Department of Endocrinology; University of Toronto, Department of Medicine, Division of Endocrinology Turner, Lucy; Ottawa Hospital Research Institute, Moher, David; Ottawa Hospital Research Institute, Ottawa Methods Centre Grimshaw, Jeremy; Ottawa Health Research Institute, Clinical Epidemiology Program
Primary Subject Heading:	Health services research
Secondary Subject Heading:	Diabetes and endocrinology, Health services research, Health policy, Evidence based practice
Keywords:	DIABETES & ENDOCRINOLOGY, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, STATISTICS & RESEARCH METHODS

Quality improvement needed in quality improvement randomized trials: Systematic review of interventions to improve care in diabetes

Noah M Ivers[1], Andrea C Tricco[2], Monica Taljaard[3], Ilana Halperin[4], Lucy Turner[5], David Moher[6], Jeremy M Grimshaw[7]

[1]Department of Family and Community Medicine, Women's College Hospital-University of Toronto, 76 Grenville Street, Toronto, Ontario, M5S1B2, Canada, *Family physician*

[2]Li Ka Shing Knowledge Institute of St Michael's Hospital, 30 Bond Street, Toronto, Ontario, M5B 1W8, Canada, *Scientist*

[3]Clinical Epidemiology Program, Ottawa Hospital Research Institute, 1053 Carling Avenue, Ottawa, Ontario, K1Y 4E9, Canada, *Scientist*

[4]Division of Endocrinology and Metabolism, Department of Medicine, University of Toronto, 200 Elizabeth St, Toronto, Ontario, M5G2C4, Canada, Endocrinology fellow

[5]Clinical Epidemiology Program, Ottawa Health Research Institute, Ottawa Hospital - General Campus, Centre for Practice-Changing Research 501 Smyth Road, Ottawa, Ontario, K1H 8L6, Canada, *Statistician*

[6]Clinical Epidemiology Program, Ottawa Health Research Institute, Ottawa Hospital - General Campus, Centre for Practice-Changing Research 501 Smyth Road, Ottawa, Ontario, K1H 8L6, Canada, *Scientist*

[7]Clinical Epidemiology Program, Ottawa Health Research Institute, Ottawa Hospital - General Campus, Centre for Practice-Changing Research 501 Smyth Road, Ottawa, Ontario, K1H 8L6, Canada, *Senior scientist*

Correspondence to: Noah Ivers, MD, Telephone: 4163236060, ext1 then 2, email: noah.ivers@utoronto.ca, fax:4163236255.

Ethical approval: Not required.

Data Sharing: Statistical code and dataset are available from the corresponding author.

Word count: 2532 (main text), 307 (abstract), 25 references, 4 tables, 2 figures.

Key words: systematic review, quality improvement, risk of bias, diabetes mellitus, randomized controlled trials

Article Summary

Article focus

- Reliable quality improvement research is needed to make decisions about initiating or scaling up quality improvement strategies.
- The number of published quality improvement trials has increased rapidly over time.
- The quality of trials published in other areas of health seem to be improving over time but the risk of bias in the quality improvement literature is uncertain

Key messages

- Nearly half of quality improvement trials for diabetes are at high risk of bias.
- The quality of quality improvement trials does not seem to be improving over time.
- Policy-makers, administrators, clinicians, and research funders must carefully scrutinize the methods used in quality improvement trials to ensure evidence-based quality improvement.

Strengths and limitations of this study

- This is the largest systematic review of risk of bias in the quality improvement literature and the only to assess for trends over time.
- The risk of bias tool does not capture all sources of methodological bias and poor reporting interferes with the assessment of many domains.
- The merits of any given trial report depends to some extent on the needs of the reader.

Abstract

Objective: Despite an increasing number of published trials of quality improvement (QI) interventions in diabetes, little is known about the risk of bias in this literature.

Design: Secondary analysis of a systematic review.

Data sources: Medline, the Cochrane Effective Practice and Organisation of Care (EPOC) database (from inception to July 2010), and references of included studies.

Eligibility criteria: Randomized trials assessing 11 predefined QI strategies or financial incentives targeting health systems, health-care professionals, or patients to improve management of adult outpatients with diabetes.

Analysis: The risk of bias (low, unclear, or high) was assessed for the 142 trials in the review across nine domains using the EPOC version of the Cochrane Risk of Bias Tool. We used Cochran-Armitage tests for trends to evaluate improvement over time

Results: There was no significant improvement over time in any of the risk of bias domains. Attrition bias (loss to follow up) was the most common source of bias, with 24 trials (17%) having high risk of bias due to incomplete outcome data. Inadequate reporting frequently hampered risk of bias assessment: allocation sequence was unclear in 82 trials (58%) and allocation concealment was unclear in 78 trials (55%). Overall, 69 trials (49%) had at least one domain with high risk of bias. There were no significant reductions in the proportions of studies that were unclear or at high risk of bias over time.

Conclusion: Nearly half of the included QI trials in this review were judged to have high risk of bias. Such trials have serious limitations that put the findings in question and therefore inhibit evidence-based QI. There is a need to limit the potential for bias when

conducting QI trials and improve the quality of reporting of QI trials so that stakeholders have adequate evidence for implementation.

Introduction

There is significant interest in quality improvement (QI) in health care, as evidenced by the rapidly increasing number of randomized clinical trials (RCTs) of QI interventions, especially in the diabetes literature.[1] RCTs can provide a foundation for making statements regarding causation, but the validity of trials varies widely; trials with adequate allocation concealment and blinding generally produce smaller effect sizes.[2] Since internal validity in QI trials is a necessary precursor for application to other settings,[3] the 'risk of bias' of the findings should be assessed to ascertain the utility of the trial results. When an RCT is deemed to have high risk of bias, the study's findings become questionable.[4]

Evaluations to assess trends in methodological quality of RCTs have been conducted in many fields of health care,[5] but no previous reviews have assessed risk of bias in QI RCTs or whether risk of bias in QI RCTs has changed over time. Recently, we conducted a systematic review and meta-regression that included 142 RCTs evaluating QI strategies to improve care for patients with diabetes.[1] In this secondary analysis of those data, we aimed to examine the risk of bias of included studies using the Cochrane Risk of Bias tool developed by the Cochrane Effective Practice and Organisation of Care (EPOC) group[6] and determine whether the proportion with high risk of bias decreased over time. We also evaluated trial and publication characteristics that might be associated with high risk of bias.

METHODS

A detailed description of the methods used for searching, screening, and abstracting the relevant data has been published[1] and is briefly summarized here.

Search strategy

Studies were identified by searching MEDLINE and the Cochrane EPOC database (up to July 2010), and screening references of included RCTs. The search strategy has been previously published[1] and is available upon request.

Study selection

RCTs examining one of eleven pre-defined QI strategies, and/or financial incentives, targeting health systems and/or healthcare professionals for the management of adult outpatients with diabetes were included. RCTs had to report at least one of the chosen process of care measures (proportion of patients taking acetylsalicylic acid, statins, anti-hypertensive medication, screened for retinopathy, screened for foot abnormalities, monitored for renal function) or intermediate outcomes (glycosylated hemoglobin levels, low-density lipoprotein cholesterol levels, diastolic and systolic blood pressure, proportion of patients with controlled hypertension, proportion of patients who quit smoking) for inclusion.

Data abstraction

A draft data abstraction form was developed and modified after a training exercise among reviewers. Two reviewers abstracted relevant data for each RCT independently.

Discrepancies were resolved by discussion or the involvement of a third reviewer. Authors of the included RCTs were contacted to obtain further information for data items requiring clarification. Journal impact factors from journal citation reports (ISI Web of Science, 2009) were obtained. When a journal's ranking was unavailable, we used the impact ranking of the open access SMImago journal and country rank database, if available.[7] This ranking is calculated using a similar formula and is strongly correlated with the journal citation impact factor.[8]

Assessing Risk of Bias

As the included trials tested QI interventions, the Cochrane EPOC Risk of Bias Tool[6] was used to assess the risk of bias in each study. The standard Cochrane Risk of Bias Tool includes an assessment of seven domains: sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting, and other. The Cochrane Handbook[9] provides instructions for making judgments about the specific domains as high, unclear, or low risk. When formulating summary assessments for each trial, classification of a study as "high risk" indicates that bias could have affected the results, while unclear risk of bias indicates that some doubt exists about the results, and low risk of bias indicates that bias is unlikely to affect the results. It has been shown empirically that studies classified as high risk using this tool are more likely to have larger effect sizes.[10]

The EPOC tool was adapted to account for the unique features of QI trials. (The guidelines for applying the Cochrane EPOC tool are summarized in Table 1.) For example, in many QI trials it is not possible to blind participants. In addition, QI trials

may require cluster-randomization to avoid contamination, but in cluster-randomized trials balance at baseline is a particular concern.[11] Therefore, the EPOC tool uses the same approach as the general Cochrane Risk of Bias Tool, but requires an assessment of bias in nine domains: sequence generation, allocation concealment, similarity of baseline measurements, similarity of baseline characteristics, incomplete data, blinding of outcome assessment, contamination, selective outcome reporting, and other. Assessment was conducted independently by a SR methodologist (ACT) and a clinician (NMI) and conflicts were resolved by discussion with an expert QI trialist (JMG).

ANALYSIS

For each risk of bias domain, the proportions of RCTs meeting the criteria for high or low or unclear risk of bias were determined. To assess for trends over time in the bias classifications, year of publication was categorized into three groups demarcated by the publication of the 2001 CONSORT statement[12] and the publication of the earlier version of the systematic review of diabetes QI interventions in 2006,[13] as we believed these may have spurred investigators to improve the quality of their trial. Therefore, we categorized year of publication as before 2002; 2002-2006; and 2007-2010.We examined each of the risk of bias domains for change over time descriptively and conducted Cochran-Armitage tests for trend for each item. Since the number of studies judged to have high risk of bias was very small for many individual domains, we grouped high and unclear risk of bias together for this test.

We estimated the proportion of QI RCTs at high risk of bias *overall*, together with 95% asymptotic confidence interval (CI). For this analysis, we created a dichotomous indicator

for each RCT based on whether or not the study was classified as high risk of bias in at least one domain. We tested for trend over time in the proportion at high risk of bias overall, hypothesizing that the proportion would decline over time. We used the same year of publication categories and conducted Cochran-Armitage tests for trend of the dichotomous indicator.

In addition, we tested for associations between high risk of bias in at least one domain and study characteristics chosen *a priori*: type of diabetes (type 1, type 2, both, unclear), type of allocation (cluster randomized, patient randomized), country (USA or Canada, UK or Western Europe, Other), type of intervention (single, multifaceted), journal impact factor, effective sample size, and year of publication using Chi-squared tests (or Fisher's exact tests, as appropriate) for categorical and Wilcoxon signed-rank tests for continuous measures. We hypothesized that each of these characteristics may be associated with studies at high risk of bias overall. All analyses were conducted in SAS Version 9.2.[14]

RESULTS

We analyzed 142 studies, with 37 (26%) published before 2002, 46 (32%) between 2002 and 2006, and 59 (42%) between 2007 and 2010. These studies evaluated the effects of QI interventions on 123,529 patients with diabetes. Trial and patient characteristics are described in Table 2. The proportions of studies judged to be at low, unclear, or high risk of bias for each domain are illustrated in Figure 1. The domains most commonly at high risk of bias were outcome reporting bias (17%) and similarity across characteristics at baseline (16%). A lack of similarity in outcome measures at baseline (10%), and lack of adequate blinding (8%) were also relatively common domains with high risk of bias.

Studies were rarely at high risk of bias due to the allocation sequence generation (4%) or allocation concealment (3%), but these domains were often unclearly reported (57% and 55% unclear, respectively). Selective outcome reporting was deemed unclear 84% of the time because published protocols were rarely available and it was often plausible that many more outcomes than those reported were measured. Table 3 indicates a lack of significant trend over time in the proportion of trials at low versus unclear or high risk of bias for any given domain.

Overall, 48.6% (69/142) of the RCTs had a high risk of bias in at least one domain (95% CI 40.4 to 56.8%). Figure 2 illustrates the rapid increase in number of QI RCTs published over time and the cumulative proportion of trials having at least one domain with high risk of bias up to a given year. In general, the line representing the proportion at high risk of bias runs parallel to the number of trials published, consistently accounting for almost half of the studies. Table 4 indicates a lack of significant trend over time in the proportion of trials with at least one domain with high risk of bias: these proportions were 46%, 44%, and 54% before 2002, between 2002 and 2006, and after 2006, respectively. Table 3 also demonstrates a lack of significant association between any of the study characteristics considered and presence of high risk of bias in at least one domain.

DISCUSSION

Main findings

Using the Cochrane EPOC Risk of Bias Tool,[6] we found that nearly half of RCTs focusing on diabetes had at least one domain at high risk of bias. The trials were most

often at high risk of bias due to inadequate follow-up of participants, a lack of similarity at baseline across outcome measures or covariates, or inadequate blinding. We also noted that the majority of RCT reports failed to include an adequate description of the allocation process (i.e., sequence generation and allocation concealment were 'unclear'). To be interpreted appropriately, RCTs must be completely and transparently reported.[15 16] Our findings indicate that greater efforts are needed to ensure both adequate reporting and methodological conduct of diabetes QI trials. For example, although blinding may be particularly difficult to accomplish in QI trials, this should be clearly reported; risk of bias could still be limited by using objective outcomes.

Comparison to literature

A systematic review focusing on cluster randomized trials found minimal improvement over time in either reporting or methodological conduct.[17] We found no evidence for a difference in the proportion of cluster randomized trials at high risk of bias compared to trials in which individuals were allocated. However, imbalance at baseline was a common source of potential bias in diabetes QI trials, possibly owing to inadequate use of restricted randomization in cluster trials.[18] Another systematic review included 35 studies covering a range of health-related fields assessing trends over time in quality criteria for RCTs.[5] Of these, 26 found improvement over time for at least one aspect of methodological quality. The domain most commonly noted to have improvement was allocation concealment, but the authors noted that this domain remained either poorly reported or inadequately performed in over half of the examined trials. We found a

similarly low proportion of studies clearly reporting adequate allocation concealment, and no evidence of improvement over time.

Previous authors have noted that QI reports may not contain enough information to inform generalization and allow for replication in different clinical settings.[19] Standards for Quality Improvement Reporting (SQUIRE) guidelines suggest that investigators conducting trials use both SQUIRE and CONSORT to inform their manuscripts.[16] Journal editors should enforce the requirements of both SQUIRE and CONSORT for QI RCTs, possibly by permitting detailed information to be posted as online appendices.

Strengths and limitations

To our knowledge, this is the largest analysis of risk of bias ever reported for health care QI RCTs and the only one to assess for trends over time. The findings are strengthened by the rigorous methods used to prepare the data for the systematic review. QI evaluations have been criticized based on numerous criteria beyond the risk of bias domains, including short duration of intervention, lack of justification for intervention design, and poor generalizability.[20 21] Some important components of methodological quality do not relate to bias (e.g. reporting of a sample size calculation). Thus, it is possible that studies at low risk of bias have important flaws with respect to methodology and/or reporting (and vice-versa), and it is possible that using other scales to assess study quality could have led to different results.[22] While the overall risk of bias assessment using the Cochrane Risk of Bias Tool has been shown to differentiate effect sizes (i.e., higher risk of bias studies usually have larger effect sizes),[10] studies at high risk of bias may still offer valuable knowledge for QI implementers. The merit of any given report

will depend on the needs of the reader, while the current analysis provides an assessment of the progress in the literature as a whole. Furthermore, we acknowledge that assigning trials with high risk of bias in a single domain a status of high risk of bias *overall* may be arguable. For this reason, we assessed trends in individual domains in addition to the summary score and also conducted a *post-hoc* sensitivity analysis that applied an empirically based rule for assigning high risk of bias overall. Previous meta-analyses have found that high risk of bias in four specific domains, namely allocation sequence generation, allocation concealment, blinding, and selective outcome reporting are each associated with greater effect size.[22-24] Our sensitivity analysis considering studies with high risk of bias in any of these four (rather than all) domains to be at high risk of bias *overall* led to the same conclusion: there has been no improvement over time in the proportion of trials at high risk of bias in this literature and no particular study characteristics were associated with high risk of bias. Another potential limitation stems from our analytical approach regarding change over time; collapsing publication year into three timeframes (pre-2002, 2002-2006, 2007-2010) and testing for trends may have limited our power. These timeframes were chosen *a priori* based on the publication of important documents that we thought might affect the conduct and reporting of these trials. We felt the assumption of linear change over time underlying the Cochran-Armitage test for trend was appropriate and in keeping with our hypotheses (e.g. high and unclear risk of bias would decrease gradually over time, while low risk of bias would increase). Risk of type 2 error is tempered by the number of tests performed; the lack of a significant p-value for trend for any level of risk of bias in any domain supports our main conclusion. Finally, this review considered only RCTs from the diabetes literature. It

would have been preferable to evaluate a random sample of all QI trials, but adequate QI electronic literature searches have yet to be developed.[25]

Implications

Published trials testing QI in diabetes are frequently at high risk of bias, producing results that may not be replicable. Clinicians must scrutinize the internal validity of the results as a first step in the process of considering the application of clinical findings for particular patients. Our findings emphasize the need for policy-makers, managers, and/or clinical-administrators seeking to implement QI interventions to apply the same process.[3] It is likely that QI investigators publishing RCTs desire for their work to have a broad impact. To help them accomplish this, research funders and journal editors can play an important role by ensuring that QI trials are reported thoroughly and transparently and are designed in a manner that limits the potential for risk of bias.

CONFLICT OF INTEREST:

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The authors declare that: (i) NM, ACT, MT, IH, LT, DM, & JG received support from Ontario Ministry of Health and Long-term Care, and the Alberta Heritage Foundation for the original systematic review, but the funding agencies had no role in the study design, collection, analysis or interpretation of data, writing of the manuscript or in the decision to submit this manuscript for publication; (ii) NM, ACT, MT, IH, LT, DM, & JG have no relationships with any companies that might have an interest in the submitted work in the previous 3 years; (iii) their spouses, partners, or children have no financial relationships that may be relevant to the submitted work; and (iv) NM, ACT, MT, IH, LT, DM, & JG have no non-financial interests that may be relevant to the submitted work.

AUTHOR CONTRIBUTIONS

NI and ACT designed and coordinated the study, participated in data collection, data analysis, data interpretation, and drafted the manuscript. MT conducted the analysis and participated in data interpretation and drafting the manuscript. IH, LT, DM, and JMG helped to design the study and write the manuscript. All authors read and approved the final manuscript.

DATA SHARING

Data are available upon request from the corresponding authour.

SOURCES OF SUPPORT

NMI holds fellowship awards from the Canadian Institutes of Health Research (CIHR) and from the Department of Family and Community Medicine, University of Toronto.

ACT holds a CIHR/Drug Safety and Effectiveness Network new investigator award.

JMG and DM both hold Canada Research Chairs.

References

1. Tricco AC, Ivers NM, Grimshaw JM, et al. Effectiveness of quality improvement strategies on the management of diabetes: a systematic review and meta-analysis. *Lancet* 2012;379:2252-2261.

2. Pildal J, Hrobjartsson A, Jorgensen KJ, et al. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. *Int J Epidemiol* 2007;36:847-857.

3. Fan E, Laupacis A, Pronovost PJ, et al. How to use an article about quality improvement. *JAMA* 2010;304:2279-2287.

4. Higgins JP, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.

5. Falagas ME, Grigori T, Ioannidou E. A systematic review of trends in the methodological quality of randomized controlled trials in various research fields. *J Clin Epidemiol* 2009;62:227-231.

6. Cochrane Effective Practice and Organisation of Care Group. Suggested risk of bias criteria for EPOC reviews, 2012. http://epoc.cochrane.org/sites/epoc.cochrane.org/files/uploads/Suggested%20risk%20of%20bias%20criteria%20for%20EPOC%20reviews.pdf

7. SCImago. SJR — SCImago Journal & Country Rank, 2007.

8. Falagas ME, Kouranos VD, Arencibia-Jorge R, et al. Comparison of SCImago journal rank indicator with journal impact factor. *FASEB J* 2008;22:2623-2628.

9. Higgins JPT, Altman DG. Assessing risk of bias in included studies. In: Higgins JPT, Green S, ed. Cochrane handbook for systematic reviews of interventions: Wiley, 2008:187-241.

10. Hartling L, Ospina M, Liang Y, et al. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. *BMJ* 2009;339:b4012.

11. Ivers NM, Halperin IJ, Barnsley J, et al. Allocation techniques for balance at baseline in cluster randomized trials: a methodological review. *Trials* 2012;13:120.

12. Moher D, Schulz KF, Altman D, et al. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* 2001;285:1987-1991

13. Shojania KG, Ranji SR, McDonald KM, et al. Effects of quality improvement strategies for type 2 diabetes on glycemic control: a meta-regression analysis. *JAMA* 2006;296:427-440.

14. SAS 9.2 [program]. Cary, NC: SAS Institute Inc., 2008.

15. Simera I, Moher D, Hirst A, et al. Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR Network. *BMC Med* 2010;8:24.

16. Davidoff F, Batalden P, Stevens D, et al. Publication guidelines for quality improvement studies in health care: evolution of the SQUIRE project. *BMJ* 2009;338:a3152.

17. Ivers NM, Taljaard M, Dixon S, et al. Impact of CONSORT extension for cluster randomised trials on quality of reporting and study methodology: review of random sample of 300 trials, 2000-8. *BMJ* 2011;343:d5886.

18. Campbell MK, Elbourne DR, Altman DG, et al. CONSORT statement: extension to cluster randomised trials. *BMJ* 2004;328(7441):702-708.

19. Michie S, Fixsen D, Grimshaw JM, et al. Specifying and reporting complex behaviour change interventions: the need for a scientific method. *Implement Sci* 2009;4:40.

20. Alexander JA, Hearld LR. What can we learn from quality improvement research? A critical review of research methods. *Med Care Res Rev* 2009;66:235-271.

21. Eccles M, Grimshaw J, Walker A, et al. Changing the behavior of healthcare professionals: the use of theory in promoting the uptake of research findings. *J Clin Epidemiol* 2005;58:107-112.

22. Armijo-Olivo S, Stiles CR, Hagen NA, et al. Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. *J Eval Clin Pract* 2012;18:12-18

23. Odgaard-Jensen J, Vist GE, Timmer A, et al. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database Syst Rev* 2011:MR000012.

24. Hempel S, Suttorp MJ, Miles JNV, et al. Empirical Evidence of Associations Between Trial Quality and Effect Size. Rockville (MD): Agency for Healthcare Research and Quality; 2011.

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

FIGURE 1: PERCENTAGE OF STUDIES JUDGED TO BE AT LOW, UNCLEAR, OR HIGH RISK OF BIAS IN EACH RISK OF BIAS DOMAIN

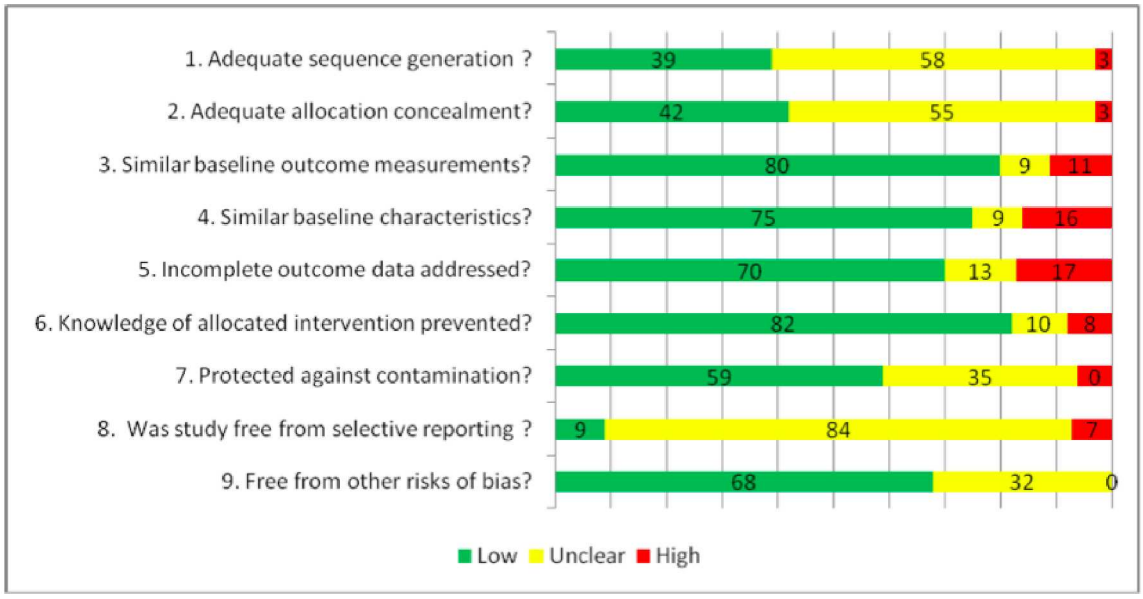


FIGURE 2: CUMULATIVE NUMBER OF DIABETES QUALITY IMPROVEMENT TRIAL PUBLICATIONS AT HIGH RISK OF BIAS IN ANY DOMAIN, 1990 TO 2010

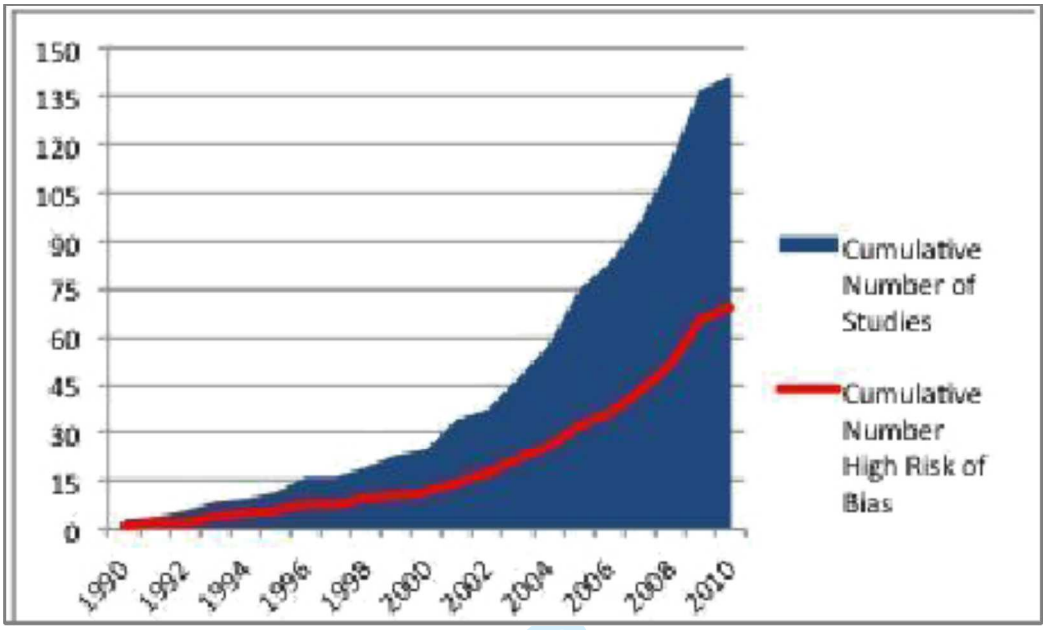


Table 1: Cochrane Effective Practice and organization of care (epoc) risk of bias assessment tool*

Risk of Bias Domain	Low Risk of Bias	High Risk of Bias	Unclear Risk of Bias
Was the allocation sequence adequately generated?	A random component in the sequence generation process is described (e.g. referring to a random number table)	Nonrandom method is used (e.g. performed by date of admission)	Not specified in the paper
Was the allocation adequately concealed?	The unit of allocation was by institution, team or professional and allocation was performed on all units at the start of the study; or if the unit of allocation was by patient or episode of care and there was some form of centralized randomization scheme, an on-site computer system or sealed opaque envelopes were used	Allocation was not adequately concealed	Not specified in the paper
Were baseline outcome measurements similar?	Performance or patient outcomes were measured prior to the intervention, and no important differences were present across study groups, or if imbalanced but appropriate adjusted analysis was performed	Important differences were present and not adjusted for in analysis	If no baseline measure of outcome
Were baseline characteristics similar?	Baseline characteristics of the study and control providers are reported and similar.	No report of characteristics in text or tables or if there are differences between control and intervention providers.	Not clear in the paper
Were incomplete outcome data adequately addressed?	Missing outcome measures were unlikely to bias the results	Missing outcome data was likely to bias the results.	Not specified in the paper
Was knowledge of the allocated interventions adequately prevented?	The authors state explicitly that the primary outcome variables were assessed blindly, or the outcomes are objective (e.g. length of hospital stay)	Outcomes were not assessed blindly and not objective.	Not specified in the paper.
Was the study free from selective outcome reporting?	There is no evidence that outcomes were selectively reported	Some important outcomes are omitted from the results	Not specified in the paper
Was the study adequately protected against contamination?	Allocation was by community, institution or practice and it is unlikely that the control group received the intervention.	It is likely that the control group received the intervention.	Communication between intervention and control professionals could have occurred
Was the study free from other risks of bias?	There is no evidence of other risk of biases		

* Adapted for ease of presentation. See http://epoc.cochrane.org/epoc-resources-review-authors for full explanation.

Table 2: Study and patient characteristics

Characteristic	Result
Patient RCTs, number (%)	94 (66.2)
Cluster RCTs, number (%)	48 (33.8)
Number of clusters, median [IQR]	29 [12, 57]
Number of patients, median [IQR]	405.3 [203, 878]
Duration of intervention months, median [IQR]	12 [8.9, 15.0]
Mean age in years, median [IQR]	59.4 [54.9, 62.9]
Percent male, median [IQR]	49.8 [41.8, 55.9]
Type of diabetes N (%)	9 (6.3)
Type 1 diabetes	80 (56.3)
Type 2 diabetes	34 (23.9)
Type 1 and 2 diabetes	19 (13.4)
Type of diabetes unclear/NR	
Number of QIs per RCT median [IQR]	2 [0, 3.5]
Administrators of patient intervention(s) N (%)	
Primary care physician	30 (21.1)
Nurse	67 (47.2)
Pharmacist	19 (13.4)
Dietician	22 (15.5)
Psychiatrist	3 (2.1)
Psychologist	2 (1.4)
Ophthalmologist	2 (1.4)
Specialist/Endocrinologist	21 (14.8)
Other	49 (34.5)
Location of study N (%)	
United States	68 (47.9)
United Kingdom	14 (9.9)
Canada	11 (7.7)
Netherlands	8 (5.6)
South Korea	7 (4.9)
Australia	6 (4.2)
Denmark	3 (2.1)
Belgium	1 (0.7)
Israel	3 (2.1)
Spain	3 (2.1)
Norway	2 (1.4)
France	2 (1.4)
Germany	2 (1.4)
Italy	2 (1.4)
Switzerland	2 (1.4)
China	2 (1.4)
Ireland	1 (0.7)
New Zealand	1 (0.7)
Thailand	1 (0.7)
Taiwan	1 (0.7)
United Arab Emirates	1 (0.7)
Mexico	1 (0.7)

Notes: † All IQRs reported as the 25th and 75th percentiles, includes investigators and community workers.
Abbreviations: RCT randomized clinical trial, N number, IQR inter-quartile range, NA not applicable, NR not reported, QI quality improvement.

Table 3: trends over time inproportions of trials classified high, unclear, or low for each risk of bias domain

RISK OF BIAS DOMAIN	Pre-2002 N=37	2002-2006 N=46	2007-2010 N=59	P-value*
Was the allocation sequence adequately generated?				**0.24**
Low	11 (30%)	19 (41%)	25 (42%)	
Unclear	24 (65%)	25 (55%)	33 (56%)	
High	2 (5%)	2 (4%)	1 (2%)	
Was the allocation adequately concealed?				**0.88**
Low	15 (40%)	20 (44%)	25 (42%)	
Unclear	21 (57%)	25 (54%)	32 (54%)	
High	1 (3%)	1 (2%)	2 (4%)	
Were baseline outcomes similar?				**0.22**
Low	31 (84%)	39 (85%)	44 (75%)	
Unclear	2 (5%)	3 (6%)	8 (13%)	
High	4 (11%)	4 (9%)	7 (12%)	
Were baseline characteristics similar?				**0.39**
Low	30 (81%)	34 (74%)	43 (73%)	
Unclear	3 (8%)	6 (13%)	3 (5%)	
High	4 (11%)	6 (13%)	13 (22%)	
Were incomplete outcome data adequately addressed?				**0.14**
Low	29 (78%)	33 (72%)	38 (64%)	
Unclear	3 (8%)	8 (17%)	7 (12%)	
High	5 (14%)	5 (11%)	14 (24%)	
Was knowledge of the allocated interventions prevented?				**0.29**
Low	32 (87%)	38 (83%)	46 (78%)	
Unclear	3 (8%)	5 (11%)	7 (12%)	
High	2 (5%)	3 (6%)	6 (10%)	
Was the study protected against contamination?				**0.55**
Low	25 (68%)	23 (50%)	35 (59%)	
Unclear	10 (27%)	21 (46%)	19 (32%)	
High	2 (5%)	2 (4%)	5 (9%)	
Was the study free from selective outcome reporting?				**0.72**
Low	3 (8%)	4 (9%)	6 (10%)	
Unclear	32 (87%)	37 (80%)	50 (85%)	
High	2 (5%)	5 (11%)	3 (5%)	
Was the study free from other risks of bias?				**0.52**
Low	27 (73%)	30 (65%)	39 (66%)	
Unclear	10 (27%)	16 (35%)	20 (34%)	
High	0	0	0	

* Cochran-Armitage test for low versus unclear or high risk of bias in each domain.

Table 4: association between study characteristics and risk of bias

Characteristic	All studies, No.	Studies in high risk of bias in at least one domain No. (%)	P-value*
Year of publication			**0.37**
Pre-2002	37	17 (46%)	
2002-2006	46	20 (44%)	
2007-2010	59	32 (54%)	
Type of diabetes			**0.11**
Type 1	9	3 (33%)	
Type 2	80	36 (45%)	
Both	34	16 (47%)	
Unclear	19	14 (74%)	
Unit of Allocation			**0.24**
Patient	94	49 (52%)	
Cluster (e.g. provider/clinic)	48	20 (42%)	
Country/Setting			**0.62**
USA or Canada	79	41 (52%)	
UK or Western Europe	40	17 (43%)	
Other	23	11 (48%)	
Journal Impact Factor			**0.87**
Greater than 3 (median)	71	34 (47.9%)	
Less than 3 (median)	71	35 (49.3%)	
Effective Sample Size			**0.87**
Greater than 154 (median)	71	35 (49.3%)	
Less than 154 (median)	71	34 (47.9%)	
Intervention Type			**0.17**
Multifaceted (featuring more than one QI strategy)	124	63 (51%)	
Single intervention	18	6 (33%)	

* Comparing proportion of studies with at least one domain at high risk of bias against studies no domains at high risk of bias. For year of publication, Cochran-Armitage test for trend was conducted. For other study characteristics, chi-squared (or Fisher's exact) tests for categorical and Wilcoxon signed-rank tests for continuous variables were used.

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	3-4
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	5
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	5-6
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	N/A
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	6-7
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	6, detailed strategy previously published
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	6-7
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	7
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	8-9
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	7-8
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	9-10

PRISMA 2009 Checklist

Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	6-9

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	23-24
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	N/A
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	9-10 (flow chart previously published)
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Previously published
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	23-24
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Risk of bias data for each study available upon request.
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	N/A
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	23
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	N/A
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	10-12
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	13-14

PRISMA 2009 Checklist

Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	15
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	2

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: **www.prisma-statement.org**.

Dear BMJ Open Editors,

It is our pleasure to submit our manuscript titled, "Quality improvement needed in quality improvement randomized trials: Systematic review of interventions to improve care in diabetes" for consideration for publication in BMJ Open.

This paper is, to our knowledge, the largest systematic review of methods used in quality improvement trials. Over the past few years, the number of randomized clinical trials evaluating interventions aiming to improve health care reported in the literature has rapidly increased. This is especially true in the diabetes literature, a condition that often acts as the exemplar for quality improvement initiatives. BMJ has played an important role in publishing quality improvement trials and in publishing research evaluating the validity of trial methodology.

The attached manuscript evaluates risk of bias using the Cochrane Risk of Bias Tool in the diabetes quality improvement literature. It is a secondary analysis of a systematic review of diabetes quality improvement interventions, which was published in the Lancet in June 2012. If the editors and/or referees deem it suitable, we would be pleased to seek permission to attach the PRISMA study flow diagram from that publication.

Our analysis in the attached manuscript demonstrates that published trials testing quality improvement interventions in diabetes are frequently at high risk of bias, producing results that may not be replicable. Worse still, the proportion at high risk of bias is not improving (decreasing) over time. The analyses in this manuscript reflect a desire to understand progression of the literature with respect to the methodological conduct of trials over time. We believe our findings in this manuscript suggest a need to carefully (re)-consider the state of the science of quality improvement as a whole given the preponderance of substandard trials in the literature.

Thank you for the opportunity to respond to the previous referee comments from BMJ. Please find the questions and our responses (italicized) below, followed by a complete version of the letter and reviews from BMJ for your reference:

Reviewer 1 –

This reviewer would encourage the authors to investigate the following improvements to their analysis:

1. You have conducted excellent work in characterizing bias in these QI trials, however can you tie the high degree of bias to QI success?
2. You have very rich and robust data abstracted from these trials. Could you use the data elements from Table 2 to conduct a meta-regression on QI success?

These suggestions reflect a desire for additional analyses, which we could pursue if the editors deem it necessary. However, we did not pursue this in our original manuscript because previous work has already shown an association between studies at high risk of bias and effect size. We refer to these studies in the manuscript and conducted a sensitivity analysis focusing on the risk of bias domains with greatest evidence for association with effect size. Furthermore, rather than testing whether risk of bias is associated with effect size, the purpose of the paper is to identify a lack of progress in the design (and reporting) of QI trials and to promote improvement in the quality of QI trials. As mentioned in the manuscript, other published papers have documented improvement over time in clinical trials, but this has never been examined in the QI field.

3. You rightly conclude CONSORT and SQUIRE guidelines should be used for forthcoming QI RCTs; please provide data on the use of CONSORT and SQUIRE in the published QI

RCTS. How was the use of CONSORT and/or SQUIRE associated with bias and QI success?
4. One contributing factor is the role of extramural funding versus unfunded, small locally run QI interventions with "randomization". Please provide data on the QI trials as to whether or not they had extramural funding: e.g., did funding play a role in the design (and thus bias) and outcome of the study?
These suggestions reflect a desire for further data abstraction from 142 published trials; this would be a daunting task even if the variables sought were well reported. However, we do not believe that the answers to the questions posed by this reviewer will be easily answered, as the variables are not likely to be readily accessible in many manuscripts. Improved reporting may reduce the number of domains marked 'unclear' but would not be expected to reduce the number of trials marked as high risk of bias. Part of the reason to encourage adherence to reporting guidelines would be to ensure that data from future trials may be more consistently extracted for methodological reviews in order to continue to monitor the state of the field. We have added a comment accordingly in the discussion.

5. Expand the discussion section to make concrete recommendations on QI trial design, evaluation, and reporting. You begin this with the SQUIRE/CONSORT statements, but could be expanded to make a statement about where the field needs to go.
The discussion section has been expanded.

Reviewer 2 –
It is reasonable to assume a time-trend in reported bias, as journals have adopted (to varying degrees) CONSORT and similar statements over the last couple of years. However, I am not so convinced of the time trend analysis as the groupings of publication year lead to loss of information and seem to suggest a significant impact of the publication of CONSORT and a previous systematic review in 2006. Unfortunately, rarely a single publication has such an impact.
We acknowledge that the time-points chosen for the analysis were somewhat arbitrary and we identify this as a potential limitation in the manuscript. Rather than trying multiple cut-points post-hoc, we planned the analysis a priori and justified the choice in the manuscript. Analyzing time as a continuous variable may have provided more power, but review of the figures and the raw numbers illustrated in the tables do not indicate any major risk for type 2 error in our analysis. Nevertheless, if the editors suggest it, we could conduct a post-hoc secondary analysis with year of publication (time) as a continuous variable.

Study characteristics are reported in Table 1. This is not very detailed and further details on the included studies have to be assessed in the companion paper published in the Lancet.
We have added to the Table a description of the QI strategies used.

"journal editors should enforce the requirements of both SQUIRE and CONSORT for QI RCTs, possibly by permitting detailed information to be posted as online appendices" should be discussed a little further, as this introduces a different issue (contextualisation, external validity) and clearly raises some questions regarding feasibility.
We have expanded the discussion section with respect to our recommendations.

Reviewer 3 –
The full list of studies included in the analysis would have been useful. I reviewed reference 1 to gain a better understanding of the studies reviewed but was unable to access the online supplementary files.
The supplementary files are available via Lancet online. We could seek permission to include the list of studies included as a supplementary file for this manuscript if the editors deem it necessary.

A table listing the QI strategies would be useful.
This has been done.

Selective outcome reporting was unclear in 85% of studies. I question the relevance of this measure of bias in a QI study.
We followed the rules set forth by the Cochrane Effective Practice and Organization of Care group in conducting this extraction. We attempted to emphasize that 'unclear' does not mean 'high' risk of bias and for this reason we did some analysis examining trends independently for 'low', 'unclear', or 'high' risk of bias.
We believe that this finding emphasizes the unique nature of QI trials, where there are often many outcomes measured since the interventions and goals are often multifaceted. It is conceivable that investigators might only report a selection of outcomes measured and that these may be more likely to be positive outcomes. For instance, consider the fact that more trials reported systolic BP than diastolic BP, when it seems inconceivable that they did not measure both. However, most studies do not publish protocols (nor do they reliably register trials in a complete manner) making it unclear whether some outcomes measured were not reported.

Question from the Editors:
We were rather surprised that this analysis was not included in the original Lancet submission. Would you be willing to tell me why that was?
The analyses in this manuscript reflect a desire to understand progression of the literature with respect to the methodological conduct of trials over time, while the Lancet review was concerned with the effect size of diabetes QI interventions. Certainly an analysis of risk of bias is standard in SRs, but an analysis of time trends in methodology/reporting is not. We believe our findings in this manuscript suggest a need to carefully (re-)consider the state of the science of quality improvement as a whole given the preponderance of substandard trials in the literature. As quality improvement trials that involve clinicians and their patients continue to increase in prevalence, we continue to believe, as mentioned by one of the referees, that the issue is relevant to a broad range of policy-makers, administrators, investigators, and clinicians.

Thank you again for considering this article at BMJ Open.

Given the ever-increasing resources dedicated to conducting such trials and the increasing interest amongst health care decision makers and stakeholders in the results of such trials, we believe the findings raise important concerns that require broad dissemination.

Sincerely,

Noah M Ivers MD CCFP PhD(c)
on behalf of the study co-authors: Andrea C Tricco, Monica Taljaard, Ilana Halperin, Lucy Turner, David Moher, and Jeremy M Grimshaw

Paper: BMJ.2012.008810

Decision: rejection

Detailed comments from the meeting:

The committee echoed that of two out of three of the reviewers who felt the work is more valuable to a specialised reader with an interest in methodology, rather than the BMJ. We were rather surprised that this analysis was not included in the original Lancet submission. Would you be willing to tell me why that was?

Please view the comments of the independent reviewers which are included at the end of this email.

You will see that one of the reviewers was more positive than we were about the paper's suitability for the BMJ.

Our main problem with the paper was that we did not think it added enough, for general readers, to what is already known about ….

We did not have any specific criticism of the design or methods.

Reviewer Comments:

Reviewer: 1

Recommendation:

Comments:
Dr. Ivers and colleagues have compiled a well-written evaluation of the quality improvement literature for diabetes by systematically reviewing 142 QI trials and providing a thorough analysis of on the role of bias. Most concerning, the author demonstrate that QI trials with a high degree of bias are rapidly increasing over the past 20 years whereby approximately 50% of the published trials today have a high degree of bias. This reviewer would encourage the authors to investigate the following improvements to their analysis:

1. You have conducted excellent work in characterizing bias in these QI trials, however can you tie the high degree of bias to QI success?

2. You have very rich and robust data abstracted from these trials. Could you use the data elements from Table 2 to conduct a meta-regression on QI success?

3. You rightly conclude CONSORT and SQUIRE guidelines should be used for forthcoming QI RCTs; please provide data on the use of CONSORT and SQUIRE in the published QI RCTS. How was the use of CONSORT and/or SQUIRE associated with bias and QI success?

4. One contributing factor is the role of extramural funding versus unfunded, small locally run QI interventions with "randomization". Please provide data on the QI trials as to whether or not they had extramural funding: e.g., did funding play a role in the design (and thus bias) and outcome of the study?

5. Expand the discussion section to make concrete recommendations on QI trial design, evaluation, and reporting. You begin this with the SQUIRE/CONSORT statements, but could be expanded to make a statement about where the field needs to go.

All in all, a excellent analysis of the QI literature and of strong interest to the BMJ readership across the spectrum from provider, administrator to researcher and editors.

Thank you

Additional Questions:
Please enter your name: Jeremiah Brown

Job Title: Assistant Professor of Health Policy and Clinical Practice

Institution: The Dartmouth Institute for Health Policy and Clinical Practice

Reimbursement for attending a symposium?: No

A fee for speaking?: No

A fee for organising education?: No

Funds for research?: Yes

Funds for a member of staff?: No

Fees for consulting?: No

Have you in the past five years been employed by an organisation that may in any way gain or lose financially from the publication of this paper?: No

Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this paper?: No

If you have any competing interests (either as indicated above or any other financial or non-financial interests) please declare them here: Agency for Healthcare Research and Quality (AHRQ) grant support for QI research

Reviewer: 2

Recommendation:

Comments:
Dear Editor and Authors,

Thank you for the opportunity to review this interesting manuscript. Please see my comments below.

Sincerely Yours,

Oliver Groene, PhD MSc MA, Lecturer in Health Services Research, London School of Hygiene and Tropical Medicine

~~~~~~

The manuscript "Quality improvement needed in quality improvement randomized trials: Systematic review of interventions to improve care in diabetes" addresses the important issue of risk of bias in the literature. This is of relevance as the literature on quality improvement interventions is increasing and a risk-of-bias assessment, according to the authors, hasn't been conducted in the field of QI trials. The authors further assess whether there is a time- trend in risk of bias, e.g. that more recent research publications are less subject to bias than earlier pieces of research. This review was prepared linked to another systematic review recently published by the author team ("Tricco AC et al. Effectiveness of quality improvement strategies on the management of diabetes: a systematic review and meta-analysis. Lancet 2012, 379: 2252-2261").

In principle, I have little to criticize: this review is clearly focused, well conducted and written concisely, and uses frameworks and tools that are widely accepted (e.g EPOC QI strategies, Cochrane-EPOC Risk of Bias tool). It is reasonable to assume a time-trend in reported bias, as journals have adopted (to varying degrees) CONSORT and similar statements over the last couple of years. However, I am not so convinced of the time trend analysis as the groupings of publication year lead to loss of information and seem to suggest a significant impact of the publication of CONSORT and a previous systematic review in 2006. Unfortunately, rarely a single publication has such an impact.

This study reports that "nearly half of the included QI trials" were judged to have a high risk of bias. This bias has not improved over time.

* Originality - does the work add enough to what is already in the published literature? If so, what does it add? If not, please cite relevant references.

The study is well conducted and adds nuances to the existing literature. Risk of bias in RCTs is a well-known fact and the authors demonstrate that this also applies to QI interventions. This in itself, I would argue, is an important message, but probably more relevant to the audience in the quality improvement field, rather than in general medicine. The authors claim that this is the only analysis that assesses trend of bias over time (page 12, line 225). If this is the case, then it would be worthwhile to pursue such an analysis on a broader literature base, not only QI trials in the field of diabetes care.

* Importance of work to general readers - does this work matter to clinicians, patients, teachers, or policymakers? Is a general journal the right place for it?

See comments above: the study has its merits, but I would think that a more specialised journal in the field of quality improvement would be more appropriate. Risk of bias in RCTs is well-known and the time trend analysis reported here is limited.

* Scientific reliability
Research Question - clearly defined and appropriately answered?
Yes, clear, sound and appropriately answered, except the general concern regarding the assessment of time trend.
~~~~~~

* Overall design of study - adequate?
Yes, except limitation in time-trend analysis

* Participants studied - adequately described and their conditions defined?
Study characteristics are reported in Table 1. This is not very detailed and further details on the included studies have to be assessed in the companion paper published in the Lancet.

* Methods - adequately described? Complies with relevant reporting standard - Eg CONSORT for randomised trials? Ethical?
Yes

* Results - answer the research question? Credible? Well presented?
Yes

* Interpretation and conclusions - warranted by and sufficiently derived from/focused on the data?
Message clear?

Yes, and the authors clearly discuss the key limitations of the paper. The conclusion that "journal editors should enforce the requirements of both SQUIRE and CONSORT for QI RCTs, possibly by permitting detailed information to be posted as online appendices" should be discussed a little further, as this introduces a different issue (contextualisation, external validity) and clearly raises some questions regarding feasibility.

* References - up to date and relevant? Any glaring omissions?
No

*Abstract/summary/key messages - reflect accurately what the paper says?
yes

Additional Questions:
Please enter your name: Oliver Groene

Job Title: 1. Lecturer Health Services Research, 2. Senior methodologist

Institution: 1. London School of Hygiene and Tropical Medicine, 2. Royal College of Surgeons

Reimbursement for attending a symposium?: No

A fee for speaking?: No

A fee for organising education?: No

Funds for research?: No

Funds for a member of staff?: No

Fees for consulting?: No

Have you in the past five years been employed by an organisation that may in any way gain or lose financially from the publication of this paper?: No

Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this paper?: No

If you have any competing interests (either as indicated above or any other financial or non-financial interests) please declare them here:

Reviewer: 3

Recommendation:

Comments:
* Originality -
The work highlights opportunities to improve the quality of QI trials

* Importance of work to general readers
I believe the work is more valuable to a specialised reader with an interest in methodology.

* Scientific reliability
The Research Question is clearly defined and appropriately answered.

* The overall design of study is adequate.
It includes recent tools developed specifically for QI studies.

* Participants studied –
The full list of studies included in the analysis would have been useful.
I reviewed reference 1. to gain a better understanding of the studies reviewed but was unable to access the online supplementary files.
A table listing the QI strategies would be useful.

* Methods –
The work complies with relevant reporting standard – SQUIRE and CONSORT.

* Results –
The research question is answered.

* Interpretation and conclusions
A general reader is likely to have different expectations of QI studies since they reflect real world conditions where loss to follow up is to be expected.
Selective outcome reporting was unclear in 85% of studies. I question the relevance of this

measure of bias in a QI study.

* References - up to date and relevant.

*Abstract/summary/key messages - reflect accurately what the paper says.

Additional Questions:
Please enter your name: Sharon Robyn O'Rourke

Job Title: Public Health Physician, Diabetes

Institution: Cairns Diabetes Centre

Reimbursement for attending a symposium?: No

A fee for speaking?: No

A fee for organising education?: No

Funds for research?: No

Funds for a member of staff?: No

Fees for consulting?: No

Have you in the past five years been employed by an organisation that may in any way gain or lose financially from the publication of this paper?: No

Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this paper?: No

If you have any competing interests (either as indicated above or any other financial or non-financial interests) please declare them here:

Quality improvement needed in quality improvement randomized trials: Systematic review of interventions to improve care in diabetes

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2013-002727.R1
Article Type:	Research
Date Submitted by the Author:	11-Mar-2013
Complete List of Authors:	Ivers, Noah; Women's College Hospital, Family and Community Medicine Tricco, Andrea; Li Ka Shing Knowledge Institute of St Michael's Hospital Taljaard, Monica; Ottawa Health Research Institute, Halperin, Ilana; Sunnybrook Health Sciences Centre, Department of Endocrinology; University of Toronto, Department of Medicine, Division of Endocrinology Turner, Lucy; Ottawa Hospital Research Institute, Moher, David; Ottawa Hospital Research Institute, Ottawa Methods Centre Grimshaw, Jeremy; Ottawa Health Research Institute, Clinical Epidemiology Program
Primary Subject Heading:	Health services research
Secondary Subject Heading:	Diabetes and endocrinology, Health services research, Health policy, Evidence based practice
Keywords:	DIABETES & ENDOCRINOLOGY, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, STATISTICS & RESEARCH METHODS

Quality improvement needed in quality improvement randomised trials: Systematic review of interventions to improve care in diabetes

Noah M Ivers[1], Andrea C Tricco[2], Monica Taljaard[3], Ilana Halperin[4], Lucy Turner[5], David Moher[6], Jeremy M Grimshaw[7]

[1]Department of Family and Community Medicine, Women's College Hospital-University of Toronto, 76 Grenville Street, Toronto, Ontario, M5S1B2, Canada, *Family physician*

[2]Li KaShing Knowledge Institute of St Michael's Hospital, 30 Bond Street, Toronto, Ontario, M5B 1W8, Canada, *Scientist*

[3]Clinical Epidemiology Program, Ottawa Hospital Research Institute, 1053 Carling Avenue, Ottawa, Ontario, K1Y 4E9, Canada, *Scientist*

[4]Division of Endocrinology and Metabolism, Department of Medicine, University of Toronto, 200 Elizabeth St, Toronto, Ontario, M5G2C4, Canada, *Endocrinology fellow*

[5]Clinical Epidemiology Program, Ottawa Health Research Institute, Ottawa Hospital - General Campus, Centre for Practice-Changing Research 501 Smyth Road, Ottawa, Ontario, K1H 8L6, Canada, *Statistician*

[6]Clinical Epidemiology Program, Ottawa Health Research Institute, Ottawa Hospital - General Campus, Centre for Practice-Changing Research 501 Smyth Road, Ottawa, Ontario, K1H 8L6, Canada, *Scientist*

[7]Clinical Epidemiology Program, Ottawa Health Research Institute, Ottawa Hospital - General Campus, Centre for Practice-Changing Research 501 Smyth Road, Ottawa, Ontario, K1H 8L6, Canada, *Senior scientist*

Correspondence to: Noah Ivers, MD, Telephone: 4163236060, ext1 then 2, email: noah.ivers@utoronto.ca, fax: 4163236255.

Ethical approval: Not required.

Data Sharing: Statistical code and dataset are available from the corresponding author.

Word count: 2797 (main text), 307 (abstract), 25 references, 4 tables, 2 figures.

Key words: systematic review, quality improvement, risk of bias, diabetes mellitus, randomised controlled trials

Article Summary

Article focus

- Reliable quality improvement research is needed to make decisions about initiating or scaling up quality improvement strategies.
- The number of published quality improvement trials has increased rapidly over time.
- The quality of trials published in other areas of health seem to be improving over time but the risk of bias in the quality improvement literature is uncertain

Key messages

- Nearly half of quality improvement trials for diabetes are at high risk of bias.
- The quality of quality improvement trials does not seem to be improving over time.
- Policy-makers, administrators, clinicians, and research funders must carefully scrutinize the methods used in quality improvement trials to ensure evidence-based quality improvement.

Strengths and limitations of this study

- This is the largest systematic review of risk of bias in the quality improvement literature and the only to assess for trends over time.
- The risk of bias tool does not capture all sources of methodological bias and poor reporting interferes with the assessment of many domains.
- The merits of any given trial report depends to some extent on the needs of the reader, such that some trials with high risk of bias may be of value for certain purposes.

Abstract

Objective: Despite increasing numbers of published trials of quality improvement (QI) interventions in diabetes, little is known about the risk of bias in this literature.

Design: Secondary analysis of a systematic review.

Data sources: Medline, the Cochrane Effective Practice and Organisation of Care (EPOC) database (from inception to July 2010), and references of included studies.

Eligibility criteria: Randomised trials assessing 11 predefined QI strategies or financial incentives targeting health systems, health-care professionals, or patients to improve management of adult outpatients with diabetes.

Analysis: Risk of bias (low, unclear, or high) was assessed for the 142 trials in the review across nine domains using the EPOC version of the Cochrane Risk of Bias Tool. We used Cochran-Armitage tests for trends to evaluate improvement over time.

Results: There was no significant improvement over time in any of the risk of bias domains. Attrition bias (loss to follow up) was the most common source of bias, with 24 trials (17%) having high risk of bias due to incomplete outcome data. Overall, 69 trials (49%) had at least one domain with high risk of bias. Inadequate reporting frequently hampered risk of bias assessment: allocation sequence was unclear in 82 trials (58%) and allocation concealment was unclear in 78 trials (55%). There were no significant reductions in the proportions of studies at high risk of bias over time, nor in the adequacy of reporting of risk of bias domains.

Conclusion: Nearly half of the included QI trials in this review were judged to have high risk of bias. Such trials have serious limitations that put the findings in question and therefore inhibit evidence-based QI. There is a need to limit the potential for bias when

conducting QI trials and improve the quality of reporting of QI trials so that stakeholders have adequate evidence for implementation.

Introduction

There is significant interest in quality improvement (QI) in health care, as evidenced by the rapidly increasing number of randomised clinical trials (RCTs) of QI interventions, especially in the diabetes literature.[1] RCTs can provide a foundation for making statements regarding causation, but the validity of trials varies widely; trials with adequate allocation concealment and blinding generally produce smaller effect sizes.[2] Since internal validity in QI trials is a necessary precursor for application to other settings,[3] the 'risk of bias' of the findings should be assessed to ascertain the utility of the trial results. When an RCT is deemed to have high risk of bias, the study's findings become questionable.[4]

Evaluations to assess trends in methodological quality of RCTs have been conducted in many fields of health care,[5] but no previous reviews have assessed risk of bias in QI RCTs or whether risk of bias in QI RCTs has changed over time. Recently, we conducted a systematic review and meta-regression that included 142 RCTs evaluating QI strategies to improve care for patients with diabetes.[1] In this secondary analysis of those data, we aimed to examine the risk of bias of included studies using the Cochrane Risk of Bias tool developed by the Cochrane Effective Practice and Organisation of Care (EPOC) group[6] and determine whether the proportion with high risk of bias decreased over time. We also evaluated trial and publication characteristics that might be associated with high risk of bias. Finally, we assessed whether the adequacy of reporting of risk of bias domains improved over time.

METHODS

A detailed description of the methods used for searching, screening, and abstracting the relevant data has been published[1] and is briefly summarized here.

Search strategy

Studies were identified by searching MEDLINE and the Cochrane EPOC database (up to July 2010), and screening references of included RCTs. The search strategy has been previously published[1] and is available upon request.

Study selection

RCTs examining one of eleven pre-defined QI strategies, and/or financial incentives, targeting health systems and/or healthcare professionals for the management of adult outpatients with diabetes were included. RCTs had to report at least one of the chosen process of care measures (proportion of patients taking acetylsalicylic acid, statins, anti-hypertensive medication, screened for retinopathy, screened for foot abnormalities, monitored for renal function) or intermediate outcomes (glycosylated haemoglobin levels, low-density lipoprotein cholesterol levels, diastolic and systolic blood pressure, proportion of patients with controlled hypertension, proportion of patients who quit smoking) for inclusion.

Data abstraction

A draft data abstraction form was developed and modified after a training exercise among reviewers. Two reviewers abstracted relevant data for each RCT independently.

Discrepancies were resolved by discussion or the involvement of a third reviewer. Authors of the included RCTs were contacted to obtain further information for data items requiring clarification. Journal impact factors from journal citation reports (ISI Web of Science, 2009) were obtained. When a journal's ranking was unavailable, we used the impact ranking of the open access SMImago journal and country rank database, if available.[7] This ranking is calculated using a similar formula and is strongly correlated with the journal citation impact factor.[8]

Assessing Risk of Bias

As the included trials tested QI interventions, the Cochrane EPOC Risk of Bias Tool[6] was used to assess the risk of bias in each study. The standard Cochrane Risk of Bias Tool includes an assessment of seven domains: sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting, and other. The Cochrane Handbook[9] provides instructions for making judgments about the specific domains as high, unclear, or low risk. When formulating summary assessments for each trial, classification of a study as "high risk" indicates that bias could have affected the results, while unclear risk of bias indicates that some doubt exists about the results, and low risk of bias indicates that bias is unlikely to affect the results. It has been shown empirically that studies classified as high risk using this tool are more likely to have larger effect sizes.[10]

The EPOC tool was adapted to account for the unique features of QI trials. (The guidelines for applying the Cochrane EPOC tool are summarized in Table 1.) For example, in many QI trials it is not possible to blind participants. In addition, QI trials

may require cluster-randomization to avoid contamination, but in cluster-randomised trials balance at baseline is a particular concern.[11] Therefore, the EPOC tool uses the same approach as the general Cochrane Risk of Bias Tool, but requires an assessment of bias in nine domains: sequence generation, allocation concealment, similarity of baseline measurements, similarity of baseline characteristics, incomplete data, blinding of outcome assessment, contamination, selective outcome reporting, and other. If a given domain is deemed 'unclear' it was inadequately reported to determine whether it meets high risk or low risk criteria. Risk of bias assessment was conducted independently by a clinician-researcher (NMI) and a systematic review methodologist (ACT) and conflicts were resolved by discussion with an expert QI trialist (JMG).

ANALYSIS

For each risk of bias domain, the proportions of RCTs meeting the criteria for high or low or unclear risk of bias were determined. To assess for trends over time in the bias classifications, year of publication was categorized into three groups demarcated by the publication of the 2001 CONSORT statement[12] and the publication of the earlier version of the systematic review of diabetes QI interventions in 2006,[13] as we believed these may have spurred investigators to improve the quality of their trial. Therefore, we categorized year of publication as before 2002; 2002-2006; and 2007-2010.We examined each of the risk of bias domains for change over time descriptively and conducted either exact or asymptotic Cochran-Armitage tests for trend for each item.

We estimated the proportion of QI RCTs at high risk of bias *overall*, together with 95% asymptotic confidence interval (CI). For this analysis, we created a dichotomous indicator

for each RCT based on whether or not the study was classified as high risk of bias in at least one domain. To assess for trends in reporting over time, we dichotomized domains as 'reported' (low or high risk of bias) and 'unreported' (unclear risk of bias). We tested for trend over time in the proportion at high risk of bias overall, hypothesizing that the proportion would decline over time. We used the same year of publication categories and conducted Cochran-Armitage tests for trend of the dichotomous indicator.

We also conducted a *post-hoc* sensitivity analysis that applied an empirically based rule for assigning high risk of bias overall. Since previous meta-analyses have found that high risk of bias in four specific domains, namely allocation sequence generation, allocation concealment, blinding, and selective outcome reporting are each associated with greater effect size,[22-24] we repeated analyses considering only studies with high risk of bias in these domains as high risk of bias *overall*.

Finally, we tested for associations between high risk of bias in at least one domain and study characteristics chosen *a priori*: type of diabetes (type 1, type 2, both, unclear), type of allocation (cluster randomised, patient randomised), country (USA or Canada, UK or Western Europe, Other), type of intervention (single, multifaceted), journal impact factor, effective sample size, and year of publication using Chi-squared tests (or Fisher's exact tests, as appropriate) for categorical and Wilcoxon signed-rank tests for continuous measures. We hypothesized that each of these characteristics may be associated with studies at high risk of bias overall.

All analyses were conducted in SAS Version 9.2.[14]

RESULTS

See Figure 1 for a study flow diagram.

We analyzed 142 studies, with 37 (26%) published before 2002, 46 (32%) between 2002 and 2006, and 59 (42%) between 2007 and 2010. These studies evaluated the effects of QI interventions on 123,529 patients with diabetes. Trial and patient characteristics are described in Table 2. The proportions of studies judged to be at low, unclear, or high risk of bias for each domain are illustrated in Figure 2. The domains most commonly at high risk of bias were outcome reporting bias (17%) and similarity across characteristics at baseline (16%). A lack of similarity in outcome measures at baseline (10%), and lack of adequate blinding (8%) were also relatively common domains with high risk of bias. Studies were rarely at high risk of bias due to the allocation sequence generation (4%) or allocation concealment (3%), but these domains were often unclearly reported (57% and 55% unclear, respectively). Selective outcome reporting was deemed unclear 84% of the time because published protocols were rarely available and it was often plausible that many more outcomes than those reported were measured. Table 3 indicates a lack of significant trend over time in the proportion of trials at high risk of bias for any given domain. Examination of Table 3 also reveals no trends over time in quality of reporting for any of the risk of bias domains.

Overall, 48.6% (69/142) of the RCTs had a high risk of bias in at least one domain (95% CI 40.4 to 56.8%). Figure 3 illustrates the rapid increase in number of QI RCTs published over time and the cumulative proportion of trials having at least one domain with high risk of bias up to a given year. In general, the line representing the proportion at high risk

of bias runs parallel to the number of trials published, consistently accounting for almost half of the studies. Table 4 indicates a lack of significant trend over time in the proportion of trials with at least one domain with high risk of bias: these proportions were 46%, 44%, and 54% before 2002, between 2002 and 2006, and after 2006, respectively. Table 4 also demonstrates a lack of significant association between any of the study characteristics considered and presence of high risk of bias in at least one domain.

The sensitivity analysis, restricting studies defined as high risk of bias overall to those with high risk of bias in one of four domains (allocation sequence generation, allocation concealment, blinding, or selective outcome reporting) also revealed no trends over time – the proportions were 19%, 20%, and 20% before 2002, between 2002 and 2006 and after 2006, respectively (p=0.86).

DISCUSSION

Main findings

Using the Cochrane EPOC Risk of Bias Tool,[6] we found that nearly half of RCTs focusing on diabetes had at least one domain at high risk of bias. The trials were most often at high risk of bias due to inadequate follow-up of participants, a lack of similarity at baseline across outcome measures or covariates, or inadequate blinding. We also noted that the majority of RCT reports failed to include an adequate description of the allocation process (i.e., sequence generation and allocation concealment were 'unclear'). To be interpreted appropriately, RCTs must be completely and transparently reported.[15,16]

Our findings indicate that greater efforts are needed to ensure both adequate reporting and methodological conduct of diabetes QI trials.

We found that poor follow-up, baseline imbalances, and blinding were the most common sources of high risk of bias. Although these domains may be difficult fully control in QI trials, methodological approaches are available to mitigate and/or explore such causes of risk of bias. For example, sensitivity analyses may be used to explore the risk of bias related to loss to follow up, and risk of baseline imbalances in QI trials may be reduced through restricted randomization techniques, especially when trials are cluster-randomized with relatively few clusters. In addition, selective outcome reporting may be limited if more QI trial protocols were registered. Finally, although blinding may be particularly difficult to accomplish in QI trials, this should be clearly reported; if outcome assessment is not blinded, risk of bias could still be limited by using objective outcomes.

Comparison to literature

A systematic review focusing on cluster randomised trials found minimal improvement over time in either reporting or methodological conduct.[17] We found no evidence for a difference in the proportion of cluster-randomised trials at high risk of bias compared to trials in which individuals were allocated. However, imbalance at baseline was a common source of potential bias in diabetes QI trials, possibly owing to inadequate use of restricted randomization in cluster trials.[18] Another systematic review included 35 studies covering a range of health-related fields assessing trends over time in quality criteria for RCTs.[5] Of these, 26 found improvement over time for at least one aspect of

methodological quality. The domain most commonly noted to have improvement was allocation concealment, but the authors noted that this domain remained either poorly reported or inadequately performed in over half of the examined trials. We found a similarly low proportion of studies clearly reporting adequate allocation concealment, and no evidence of improvement over time.

Previous authors have noted that QI reports may not contain enough information to inform generalization and allow for replication in different clinical settings.[19] Standards for Quality Improvement Reporting (SQUIRE) guidelines suggest that investigators conducting trials use both SQUIRE and CONSORT to inform their manuscripts.[16] Journal editors should enforce the requirements of both SQUIRE and CONSORT for QI RCTs, possibly by permitting detailed information to be posted as online appendices. Although it might seem onerous to force investigators to address all items in SQUIRE and CONSORT, the risks of poor reporting are substantial. Inadequate description of context could omit essential pre-conditions or important effect modifiers for a successful QI program, while incomplete description of the program itself might lead to failure due to partial implementation.

Strengths and limitations

To our knowledge, this is the largest analysis of risk of bias ever reported for health care QI RCTs and the only one to assess for trends over time. The findings are strengthened by the rigorous methods used to prepare the data for the systematic review.

QI evaluations have been criticized based on numerous criteria beyond the risk of bias domains, including short duration of intervention, lack of justification for intervention design, and poor generalizability.[2021] Some important components of methodological quality do not relate to bias (e.g. reporting of a sample size calculation). Thus, it is possible that studies at low risk of bias have important flaws with respect to methodology and/or reporting (and vice-versa), and it is possible that using other scales to assess study quality could have led to different results.[22] While the overall risk of bias assessment using the Cochrane Risk of Bias Tool has been shown to differentiate effect sizes (i.e., higher risk of bias studies usually have larger effect sizes),[10] studies at high risk of bias may still offer valuable knowledge for QI implementers. The merit of any given report will depend on the needs of the reader, while the current analysis provides an assessment of the progress in the literature as a whole.

Furthermore, we acknowledge that assigning trials with high risk of bias in a single domain a status of high risk of bias *overall* may be arguable. Nevertheless, our sensitivity analysis led to the same conclusion: there has been no improvement over time in the proportion of trials at high risk of bias in this literature and no particular study characteristics were associated with high risk of bias.

Another potential limitation stems from our analytical approach regarding change over time; collapsing publication year into three timeframes (pre-2002, 2002-2006, 2007-2010) and testing for trends may have limited our power. These timeframes were chosen *a priori* based on the publication of important documents that we thought might affect the conduct and reporting of these trials. We felt the assumption of linear change over time

underlying the Cochran-Armitage test for trend was appropriate and in keeping with our hypotheses (e.g. high and unclear risk of bias would decrease gradually over time, while low risk of bias would increase). Risk of type 2 error is tempered by the number of tests performed; the lack of a significant p-value for trend for any level of risk of bias in any domain supports our main conclusion. Finally, this review considered only RCTs from the diabetes literature. It would have been preferable to evaluate a random sample of all QI trials, but adequate QI electronic literature searches have yet to be developed.[25]

Implications

Published trials testing QI in diabetes are frequently at high risk of bias, producing results that may not be replicable. Clinicians must scrutinize the internal validity of the results as a first step in the process of considering the application of clinical findings for particular patients. Our findings emphasize the need for policy-makers, managers, and/or clinical-administrators seeking to implement QI interventions to apply the same process.[3] It is likely that QI investigators publishing RCTs desire for their work to have a broad impact. To help them accomplish this, research funders and journal editors can play an important role by ensuring that QI trials are reported thoroughly and transparently and are designed in a manner that limits the potential for risk of bias.

CONFLICT OF INTEREST:

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The authors declare that: (i) NM, ACT, MT, IH, LT, DM, & JG received support from Ontario Ministry of Health and Long-term Care, and the Alberta Heritage Foundation for the original systematic review, but the funding agencies had no role in the study design, collection, analysis or interpretation of data, writing of the manuscript or in the decision to submit this manuscript for publication; (ii) NM, ACT, MT, IH, LT, DM, & JG have no relationships with any companies that might have an interest in the submitted work in the previous 3 years; (iii) their spouses, partners, or children have no financial relationships that may be relevant to the submitted work; and (iv) NM, ACT, MT, IH, LT, DM, & JG have no non-financial interests that may be relevant to the submitted work.

AUTHOR CONTRIBUTIONS

NI and ACT designed and coordinated the study, participated in data collection, data analysis, data interpretation, and drafted the manuscript. MT conducted the analysis and participated in data interpretation and drafting the manuscript. IH, LT, DM, and JMG helped to design the study and write the manuscript. All authors read and approved the final manuscript.

DATA SHARING

Data are available upon request from the corresponding authour.

SOURCES OF SUPPORT

NMI holds fellowship awards from the Canadian Institutes of Health Research (CIHR) and from the Department of Family and Community Medicine, University of Toronto.

ACT holds a CIHR/Drug Safety and Effectiveness Network new investigator award.

JMG and DM both hold Canada Research Chairs.

References

1. Tricco AC, Ivers NM, Grimshaw JM, et al. Effectiveness of quality improvement strategies on the management of diabetes: a systematic review and meta-analysis. *Lancet* 2012;379:2252-2261.

2. Pildal J, Hrobjartsson A, Jorgensen KJ, et al. Impact of allocation concealment on conclusions drawn from meta-analyses of randomised trials. *Int J Epidemiol* 2007;36:847-857.

3. Fan E, Laupacis A, Pronovost PJ, et al. How to use an article about quality improvement. *JAMA* 2010;304:2279-2287.

4. Higgins JP, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.

5. Falagas ME, Grigori T, Ioannidou E. A systematic review of trends in the methodological quality of randomised controlled trials in various research fields. *J Clin Epidemiol* 2009;62:227-231.

6. Cochrane Effective Practice and Organisation of Care Group. Suggested risk of bias criteria for EPOC reviews, 2012. http://epoc.cochrane.org/sites/epoc.cochrane.org/files/uploads/Suggested%20risk%20of%20bias%20criteria%20for%20EPOC%20reviews.pdf

7. SCImago. SJR — SCImago Journal & Country Rank, 2007.

8. Falagas ME, Kouranos VD, Arencibia-Jorge R, et al. Comparison of SCImago journal rank indicator with journal impact factor. *FASEB J* 2008;22:2623-2628.

9. Higgins JPT, Altman DG. Assessing risk of bias in included studies. In: Higgins JPT, Green S, ed. Cochrane handbook for systematic reviews of interventions: Wiley, 2008:187-241.

10. Hartling L, Ospina M, Liang Y, et al. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. *BMJ* 2009;339:b4012.

11. Ivers NM, Halperin IJ, Barnsley J, et al. Allocation techniques for balance at baseline in cluster randomised trials: a methodological review. *Trials* 2012;13:120.

12. Moher D, Schulz KF, Altman D, et al. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *JAMA* 2001;285:1987-1991

13. Shojania KG, Ranji SR, McDonald KM, et al. Effects of quality improvement strategies for type 2 diabetes on glycemic control: a meta-regression analysis. *JAMA* 2006;296:427-440.

14. SAS 9.2 [program]. Cary, NC: SAS Institute Inc., 2008.

15. Simera I, Moher D, Hirst A, et al. Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR Network. *BMC Med* 2010;8:24.

16. Davidoff F, Batalden P, Stevens D, et al. Publication guidelines for quality improvement studies in health care: evolution of the SQUIRE project. *BMJ* 2009;338:a3152.

17. Ivers NM, Taljaard M, Dixon S, et al. Impact of CONSORT extension for cluster randomised trials on quality of reporting and study methodology: review of random sample of 300 trials, 2000-8. *BMJ* 2011;343:d5886.

18. Campbell MK, Elbourne DR, Altman DG, et al. CONSORT statement: extension to cluster randomised trials. *BMJ* 2004;328(7441):702-708.

19. Michie S, Fixsen D, Grimshaw JM, et al. Specifying and reporting complex behaviour change interventions: the need for a scientific method. *Implement Sci* 2009;4:40.

20. Alexander JA, Hearld LR. What can we learn from quality improvement research? A critical review of research methods. *Med Care Res Rev* 2009;66:235-271.

21. Eccles M, Grimshaw J, Walker A, et al. Changing the behavior of healthcare professionals: the use of theory in promoting the uptake of research findings. *J Clin Epidemiol* 2005;58:107-112.

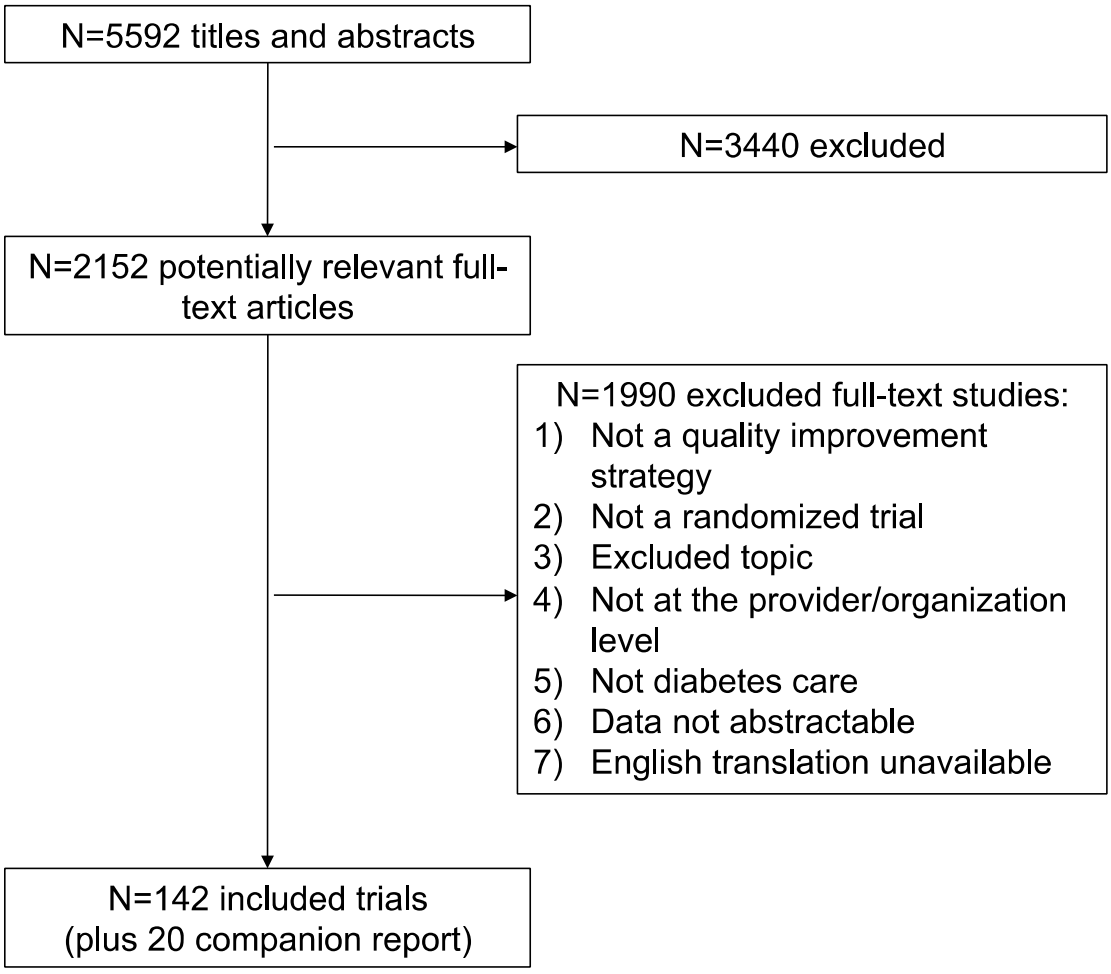
22. Armijo-Olivo S, Stiles CR, Hagen NA, et al. Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. *J Eval Clin Pract* 2012;18:12-18

23. Odgaard-Jensen J, Vist GE, Timmer A, et al. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database Syst Rev* 2011:MR000012.

24. Hempel S, Suttorp MJ, Miles JNV, et al. Empirical Evidence of Associations Between Trial Quality and Effect Size. Rockville (MD): Agency for Healthcare Research and Quality; 2011.

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

Figure 1: Study flow diagram



403

Figure 2: percentage of studies judged to be at low, unclear, or high risk of bias in each risk of bias domain

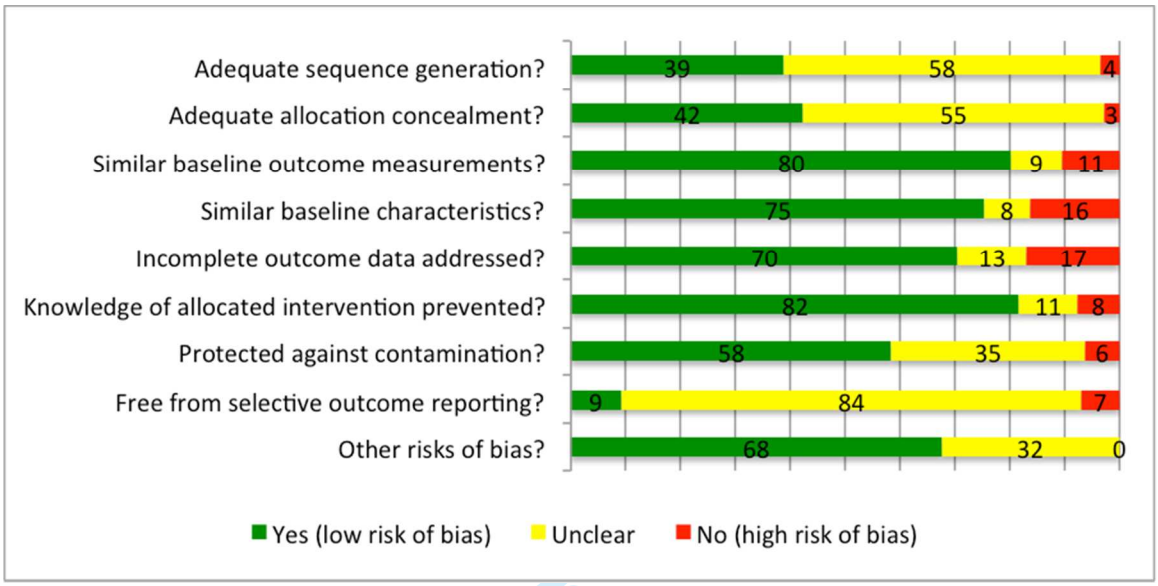


Figure 3: cumulative number of diabetes quality improvement trial publications at high risk of bias in any domain, 1990 to 2010.

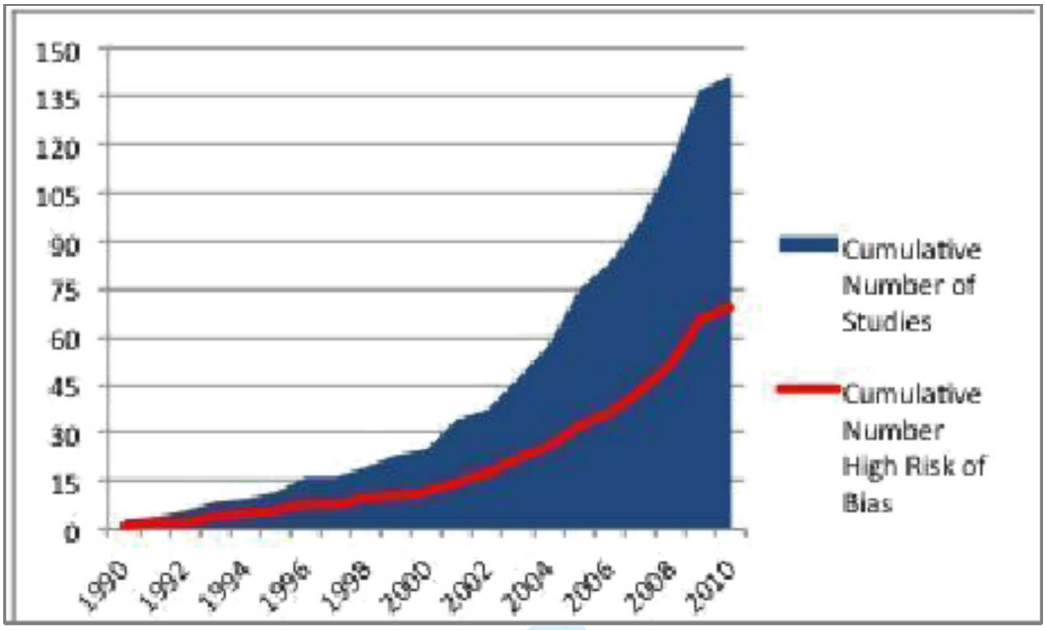


Table 1: Cochrane Effective Practice and organization of care (epoc) risk of bias assessment tool*

Risk of Bias Domain	Low Risk of Bias	High Risk of Bias	Unclear Risk of Bias
Was the allocation sequence adequately generated?	A random component in the sequence generation process is described (e.g. referring to a random number table)	Nonrandom method is used (e.g. performed by date of admission)	Not specified in the paper
Was the allocation adequately concealed?	The unit of allocation was by institution, team or professional and allocation was performed on all units at the start of the study; or if the unit of allocation was by patient or episode of care and there was some form of centralized randomization scheme, an on-site computer system or sealed opaque envelopes were used	Allocation was not adequately concealed	Not specified in the paper
Were baseline outcome measurements similar?	Performance or patient outcomes were measured prior to the intervention, and no important differences were present across study groups, or if imbalanced but appropriate adjusted analysis was performed	Important differences were present and not adjusted for in analysis	If no baseline measure of outcome
Were baseline characteristics similar?	Baseline characteristics of the study and control providers are reported and similar.	No report of characteristics in text or tables or if there are differences between control and intervention providers.	Not clear in the paper
Were incomplete outcome data adequately addressed?	Missing outcome measures were unlikely to bias the results	Missing outcome data was likely to bias the results.	Not specified in the paper
Was knowledge of the allocated interventions adequately prevented?	The authors state explicitly that the primary outcome variables were assessed blindly, or the outcomes are objective (e.g. length of hospital stay)	Outcomes were not assessed blindly and not objective.	Not specified in the paper.
Was the study free from selective outcome reporting?	There is no evidence that outcomes were selectively reported	Some important outcomes are omitted from the results	Not specified in the paper
Was the study adequately protected against contamination?	Allocation was by community, institution or practice and it is unlikely that the control group received the intervention.	It is likely that the control group received the intervention.	Communication between intervention and control professionals could have occurred
Was the study free from other risks of bias?	There is no evidence of other risk of biases		

* Adapted for ease of presentation. See http://epoc.cochrane.org/epoc-resources-review-authors for full explanation.

Table 2: Study and patient characteristics

Characteristic	Result
Patient RCTs, number (%)	94 (66.2)
Cluster RCTs, number (%)	48 (33.8)
Number of clusters, median [IQR]	29 [12, 57]
Numberof patients, median [IQR]	405.3 [203, 878]
Duration of intervention months, median [IQR]	12 [8.9, 15.0]
Mean age in years, median [IQR]	59.4 [54.9, 62.9]
Percent male, median [IQR]	49.8 [41.8, 55.9]
Type of diabetes N (%)	9 (6.3)
Type 1 diabetes	80 (56.3)
Type 2 diabetes	34 (23.9)
Type 1 and 2 diabetes	19 (13.4)
Type of diabetes unclear/NR	
Number of QIs per RCT median [IQR]	2 [0, 3.5]
Administrators of patient intervention(s) N (%)	
Primary care physician	30 (21.1)
Nurse	67 (47.2)
Pharmacist	19 (13.4)
Dietician	22 (15.5)
Psychiatrist	3 (2.1)
Psychologist	2 (1.4)
Ophthalmologist	2 (1.4)
Specialist/Endocrinologist	21 (14.8)
Other	49 (34.5)
Location of study N (%)	
United States	68 (47.9)
United Kingdom	14 (9.9)
Canada	11 (7.7)
Netherlands	8 (5.6)
South Korea	7 (4.9)
Australia	6 (4.2)
Denmark	3 (2.1)
Belgium	1 (0.7)
Israel	3 (2.1)
Spain	3 (2.1)
Norway	2 (1.4)
France	2 (1.4)
Germany	2 (1.4)
Italy	2 (1.4)
Switzerland	2 (1.4)
China	2 (1.4)
Ireland	1 (0.7)
New Zealand	1 (0.7)
Thailand	1 (0.7)
Taiwan	1 (0.7)
United Arab Emirates	1 (0.7)
Mexico	1 (0.7)

Notes: † All IQRs reported as the 25th and 75th percentiles, includes investigators and community workers.
Abbreviations: RCT randomised clinical trial, N number, IQR inter-quartile range, NA not applicable, NR not reported, QI quality improvement.

Table 3: Trends over time in proportions of trials classified high, unclear, or low for each risk of bias domain

RISK OF BIAS DOMAIN	Pre-2002 N=37	2002-2006 N=46	2007-2010 N=59	P-value*	P-value^
Was the allocation sequence adequately generated?				**0.41**	**0.43**
Low	11 (30%)	19 (41%)	25 (42%)		
Unclear	24 (65%)	25 (55%)	33 (56%)		
High	2 (5%)	2 (4%)	1 (2%)		
Was the allocation adequately concealed?				**1.00**	**0.82**
Low	15 (40%)	20 (44%)	25 (42%)		
Unclear	21 (57%)	25 (54%)	32 (54%)		
High	1 (3%)	1 (2%)	2 (4%)		
Were baseline outcomes similar?				**0.87**	**0.20**
Low	31 (84%)	39 (85%)	44 (75%)		
Unclear	2 (5%)	3 (6%)	8 (13%)		
High	4 (11%)	4 (9%)	7 (12%)		
Were baseline characteristics similar?				**0.16**	**0.57**
Low	30 (81%)	34 (74%)	43 (73%)		
Unclear	3 (8%)	6 (13%)	3 (5%)		
High	4 (11%)	6 (13%)	13 (22%)		
Were incomplete outcome data adequately addressed?				**0.17**	**0.70**
Low	29 (78%)	33 (72%)	38 (64%)		
Unclear	3 (8%)	8 (17%)	7 (12%)		
High	5 (14%)	5 (11%)	14 (24%)		
Was knowledge of the allocated interventions prevented?				**0.44**	**0.61**
Low	32 (87%)	38 (83%)	46 (78%)		
Unclear	3 (8%)	5 (11%)	7 (12%)		
High	2 (5%)	3 (6%)	6 (10%)		
Was the study protected against contamination?				**0.54**	**0.78**
Low	25 (68%)	23 (50%)	35 (59%)		
Unclear	10 (27%)	21 (46%)	19 (32%)		
High	2 (5%)	2 (4%)	5 (9%)		
Was the study free from selective outcome reporting?				**0.84**	**1.00**
Low	3 (8%)	4 (9%)	6 (10%)		
Unclear	32 (87%)	37 (80%)	50 (85%)		
High	2 (5%)	5 (11%)	3 (5%)		
Was the study free from other risks of bias?				**0.58**	**0.58**
Low	27 (73%)	30 (65%)	39 (66%)		
Unclear	10 (27%)	16 (35%)	20 (34%)		
High	0	0	0		

* Exact Cochran-Armitage test for high versus low or unclear risk of bias in each domain except the last domain which was analyzed as low versus high or unclear due to absence of studies with high risk of bias.
^ Exact Cochran-Armitage test for reported (high or low risk of bias) or unreported (unclear risk of bias) in each domain.

Table 4: association between study characteristics and risk of bias

Characteristic	All studies, No.	Studies in high risk of bias in at least one domain No. (%)	P-value*
Year of publication			**0.37**
Pre-2002	37	17 (46%)	
2002-2006	46	20 (44%)	
2007-2010	59	32 (54%)	
Type of diabetes			**0.11**
Type 1	9	3 (33%)	
Type 2	80	36 (45%)	
Both	34	16 (47%)	
Unclear	19	14 (74%)	
Unit of Allocation			**0.24**
Patient	94	49 (52%)	
Cluster (e.g. provider/clinic)	48	20 (42%)	
Country/Setting			**0.62**
USA or Canada	79	41 (52%)	
UK or Western Europe	40	17 (43%)	
Other	23	11 (48%)	
Journal Impact Factor			**0.87**
Greater than 3 (median)	71	34 (47.9%)	
Less than 3 (median)	71	35 (49.3%)	
Effective Sample Size			**0.87**
Greater than 154 (median)	71	35 (49.3%)	
Less than 154 (median)	71	34 (47.9%)	
Intervention Type			**0.17**
Multifaceted (featuring more than one QI strategy)	124	63 (51%)	
Single intervention	18	6 (33%)	

* Comparing proportion of studies with at least one domain at high risk of bias against studies no domains at high risk of bias. For year of publication, Cochran-Armitage test for trend was conducted. For other study characteristics, chi-squared (or Fisher's exact) tests for categorical and Wilcoxon signed-rank tests for continuous variables were used

Quality improvement needed in quality improvement r~~andomized~~andomised trials: Systematic review of interventions to improve care in diabetes

Noah M Ivers[1], Andrea C Tricco[2], Monica Taljaard[3], Ilana Halperin[4], Lucy Turner[5], David Moher[6], Jeremy M Grimshaw[7]

[1]Department of Family and Community Medicine, Women's College Hospital-University of Toronto, 76 Grenville Street, Toronto, Ontario, M5S1B2, Canada, *Family physician*

[2]Li KaShing Knowledge Institute of St Michael's Hospital, 30 Bond Street, Toronto, Ontario, M5B 1W8, Canada, *Scientist*

[3]Clinical Epidemiology Program, Ottawa Hospital Research Institute, 1053 Carling Avenue, Ottawa, Ontario, K1Y 4E9, Canada, *Scientist*

[4]Division of Endocrinology and Metabolism, Department of Medicine, University of Toronto, 200 Elizabeth St, Toronto, Ontario, M5G2C4, Canada, Endocrinology fellow

[5]Clinical Epidemiology Program, Ottawa Health Research Institute, Ottawa Hospital - General Campus, Centre for Practice-Changing Research 501 Smyth Road, Ottawa, Ontario, K1H 8L6, Canada, *Statistician*

[6]Clinical Epidemiology Program, Ottawa Health Research Institute, Ottawa Hospital - General Campus, Centre for Practice-Changing Research 501 Smyth Road, Ottawa, Ontario, K1H 8L6, Canada, *Scientist*

[7]Clinical Epidemiology Program, Ottawa Health Research Institute, Ottawa Hospital - General Campus, Centre for Practice-Changing Research 501 Smyth Road, Ottawa, Ontario, K1H 8L6, Canada, *Senior scientist*

Correspondence to: Noah Ivers, MD, Telephone: 4163236060, ext1 then 2, email: noah.ivers@utoronto.ca, fax:4163236255.

Ethical approval: Not required.

Data Sharing: Statistical code and dataset are available from the corresponding author.

Word count: 2532 (main text), 307 (abstract), 25 references, 4 tables, 2 figures.

Key words: systematic review, quality improvement, risk of bias, diabetes mellitus, ~~randomized~~randomised controlled trials

Article Summary

Article focus

- Reliable quality improvement research is needed to make decisions about initiating or scaling up quality improvement strategies.
- The number of published quality improvement trials has increased rapidly over time.
- The quality of trials published in other areas of health seem to be improving over time but the risk of bias in the quality improvement literature is uncertain

Key messages

- Nearly half of quality improvement trials for diabetes are at high risk of bias.
- The quality of quality improvement trials does not seem to be improving over time.
- Policy-makers, administrators, clinicians, and research funders must carefully scrutinize the methods used in quality improvement trials to ensure evidence-based quality improvement.

Strengths and limitations of this study

- This is the largest systematic review of risk of bias in the quality improvement literature and the only to assess for trends over time.
- The risk of bias tool does not capture all sources of methodological bias and poor reporting interferes with the assessment of many domains.
- The merits of any given trial report depends to some extent on the needs of the reader, such that some trials with high risk of bias may be of value for certain purposes.

Abstract

Objective: Despite an increasing number of published trials of quality improvement (QI) interventions in diabetes, little is known about the risk of bias in this literature.

Design: Secondary analysis of a systematic review.

Data sources: Medline, the Cochrane Effective Practice and Organisation of Care (EPOC) database (from inception to July 2010), and references of included studies.

Eligibility criteria: R~~andomized~~andomised trials assessing 11 predefined QI strategies or financial incentives targeting health systems, health-care professionals, or patients to improve management of adult outpatients with diabetes.

Analysis: The risk of bias (low, unclear, or high) was assessed for the 142 trials in the review across nine domains using the EPOC version of the Cochrane Risk of Bias Tool. We used Cochran-Armitage tests for trends to evaluate improvement over time.

Results: There was no significant improvement over time in any of the risk of bias domains. Attrition bias (loss to follow up) was the most common source of bias, with 24 trials (17%) having high risk of bias due to incomplete outcome data. Inadequate reporting frequently hampered risk of bias assessment: allocation sequence was unclear in 82 trials (58%) and allocation concealment was unclear in 78 trials (55%). Overall, 69 trials (49%) had at least one domain with high risk of bias. There were no significant reductions in the proportions of studies that ~~were~~ at ~~unclear or at~~ high risk of bias ~~-~~over time, nor in the adequacy of reporting of risk of bias domains.

Conclusion: Nearly half of the included QI trials in this review were judged to have high risk of bias. Such trials have serious limitations that put the findings in question and therefore inhibit evidence-based QI. There is a need to limit the potential for bias when

conducting QI trials and improve the quality of reporting of QI trials so that stakeholders have adequate evidence for implementation.

Introduction

There is significant interest in quality improvement (QI) in health care, as evidenced by the rapidly increasing number of ~~randomized~~randomised clinical trials (RCTs) of QI interventions, especially in the diabetes literature.[1] RCTs can provide a foundation for making statements regarding causation, but the validity of trials varies widely; trials with adequate allocation concealment and blinding generally produce smaller effect sizes.[2] Since internal validity in QI trials is a necessary precursor for application to other settings,[3] the 'risk of bias' of the findings should be assessed to ascertain the utility of the trial results. When an RCT is deemed to have high risk of bias, the study's findings become questionable.[4]

Evaluations to assess trends in methodological quality of RCTs have been conducted in many fields of health care,[5] but no previous reviews have assessed risk of bias in QI RCTs or whether risk of bias in QI RCTs has changed over time. Recently, we conducted a systematic review and meta-regression that included 142 RCTs evaluating QI strategies to improve care for patients with diabetes.[1] In this secondary analysis of those data, we aimed to examine the risk of bias of included studies using the Cochrane Risk of Bias tool developed by the Cochrane Effective Practice and Organisation of Care (EPOC) group[6] and determine whether the proportion with high risk of bias decreased over time. We also evaluated trial and publication characteristics that might be associated with high risk of bias. Finally, we assessed whether the adequacy of reporting of risk of bias domains improved over time.

METHODS

A detailed description of the methods used for searching, screening, and abstracting the relevant data has been published[1] and is briefly summarized here.

Search strategy

Studies were identified by searching MEDLINE and the Cochrane EPOC database (up to July 2010), and screening references of included RCTs. The search strategy has been previously published[1] and is available upon request.

Study selection

RCTs examining one of eleven pre-defined QI strategies, and/or financial incentives, targeting health systems and/or healthcare professionals for the management of adult outpatients with diabetes were included. RCTs had to report at least one of the chosen process of care measures (proportion of patients taking acetylsalicylic acid, statins, anti-hypertensive medication, screened for retinopathy, screened for foot abnormalities, monitored for renal function) or intermediate outcomes (glycosylated haemoglobin levels, low-density lipoprotein cholesterol levels, diastolic and systolic blood pressure, proportion of patients with controlled hypertension, proportion of patients who quit smoking) for inclusion.

Data abstraction

A draft data abstraction form was developed and modified after a training exercise among reviewers. Two reviewers abstracted relevant data for each RCT independently.

Discrepancies were resolved by discussion or the involvement of a third reviewer. Authors of the included RCTs were contacted to obtain further information for data items requiring clarification. Journal impact factors from journal citation reports (ISI Web of Science, 2009) were obtained. When a journal's ranking was unavailable, we used the impact ranking of the open access SMImago journal and country rank database, if available.[7] This ranking is calculated using a similar formula and is strongly correlated with the journal citation impact factor.[8]

Assessing Risk of Bias

As the included trials tested QI interventions, the Cochrane EPOC Risk of Bias Tool[6] was used to assess the risk of bias in each study. The standard Cochrane Risk of Bias Tool includes an assessment of seven domains: sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting, and other. The Cochrane Handbook[9] provides instructions for making judgments about the specific domains as high, unclear, or low risk. When formulating summary assessments for each trial, classification of a study as "high risk" indicates that bias could have affected the results, while unclear risk of bias indicates that some doubt exists about the results, and low risk of bias indicates that bias is unlikely to affect the results. It has been shown empirically that studies classified as high risk using this tool are more likely to have larger effect sizes.[10]

The EPOC tool was adapted to account for the unique features of QI trials. (The guidelines for applying the Cochrane EPOC tool are summarized in Table 1.) For example, in many QI trials it is not possible to blind participants. In addition, QI trials

may require cluster-randomization to avoid contamination, but in cluster-randomizedandomised trials balance at baseline is a particular concern.[11] Therefore, the EPOC tool uses the same approach as the general Cochrane Risk of Bias Tool, but requires an assessment of bias in nine domains: sequence generation, allocation concealment, similarity of baseline measurements, similarity of baseline characteristics, incomplete data, blinding of outcome assessment, contamination, selective outcome reporting, and other. If a given domain is deemed 'unclear' it was inadequately reported to determine whether it meets high risk or low risk criteria. Risk of bias aAssessment was conducted independently by a a clinician-researcher (NMI) and a systematic reviewSR methodologist (ACT) and a clinician (NMI) and conflicts were resolved by discussion with an expert QI trialist (JMG).

ANALYSIS

For each risk of bias domain, the proportions of RCTs meeting the criteria for high or low or unclear risk of bias were determined. To assess for trends over time in the bias classifications, year of publication was categorized into three groups demarcated by the publication of the 2001 CONSORT statement[12] and the publication of the earlier version of the systematic review of diabetes QI interventions in 2006,[13] as we believed these may have spurred investigators to improve the quality of their trial. Therefore, we categorized year of publication as before 2002; 2002-2006; and 2007-2010.We examined each of the risk of bias domains for change over time descriptively and conducted either exact or asymptotic Cochran-Armitage tests for trend for each item. Since the number of studies

~~judged to have high risk of bias was very small for many individual domains, we grouped high and unclear risk of bias together for this test.~~

We estimated the proportion of QI RCTs at high risk of bias *overall*, together with 95% asymptotic confidence interval (CI). For this analysis, we created a dichotomous indicator for each RCT based on whether or not the study was classified as high risk of bias in at least one domain. To assess for trends in reporting over time, we dichotomized domains as 'reported' (low or high risk of bias) and 'unreported' (unclear risk of bias). We tested for trend over time in the proportion at high risk of bias overall, hypothesizing that the proportion would decline over time. We used the same year of publication categories and conducted Cochran-Armitage tests for trend of the dichotomous indicator.

~~For this reaso~~We also conducted a *post-hoc* sensitivity analysis that applied an empirically based rule for assigning high risk of bias overall. ~~n, we assessed trends in individual domains in addition to the summary score and also conducted a~~Since ~~*post-hoc* sensitivity analysis that applied an empirically based rule for assigning high risk of bias overall. p~~Previous meta-analyses have found that high risk of bias in four specific domains, namely allocation sequence generation, allocation concealment, blinding, and selective outcome reporting are each associated with greater effect size,[22-24] we repeated analyses considering only studies with high risk of bias in these domains as high risk of bias *overall*.

~~In addition~~Finally, we tested for associations between high risk of bias in at least one domain and study characteristics chosen *a priori*: type of diabetes (type 1, type 2, both, unclear), type of allocation (cluster r~~andomized~~andomised, patient

Field Code Changed

Formatted: Font: Italic

randomizedandomised), country (USA or Canada, UK or Western Europe, Other), type of intervention (single, multifaceted), journal impact factor, effective sample size, and year of publication using Chi-squared tests (or Fisher's exact tests, as appropriate) for categorical and Wilcoxon signed-rank tests for continuous measures. We hypothesized that each of these characteristics may be associated with studies at high risk of bias overall.

All analyses were conducted in SAS Version 9.2.[14]

RESULTS

See Figure 1 for a study flow diagram.

Formatted: Space After: 12 pt

We analyzed 142 studies, with 37 (26%) published before 2002, 46 (32%) between 2002 and 2006, and 59 (42%) between 2007 and 2010. These studies evaluated the effects of QI interventions on 123,529 patients with diabetes. Trial and patient characteristics are described in Table 2. The proportions of studies judged to be at low, unclear, or high risk of bias for each domain are illustrated in Figure 21. The domains most commonly at high risk of bias were outcome reporting bias (17%) and similarity across characteristics at baseline (16%). A lack of similarity in outcome measures at baseline (10%), and lack of adequate blinding (8%) were also relatively common domains with high risk of bias. Studies were rarely at high risk of bias due to the allocation sequence generation (4%) or allocation concealment (3%), but these domains were often unclearly reported (57% and 55% unclear, respectively). Selective outcome reporting was deemed unclear 84% of the time because published protocols were rarely available and it was often plausible that

many more outcomes than those reported were measured. Table 3 indicates a lack of significant trend over time in the proportion of trials at ~~low versus unclear or~~ high risk of bias for any given domain. Examination of Table 3 also reveals no trends over time in quality of reporting for any of the risk of bias domains.

Overall, 48.6% (69/142) of the RCTs had a high risk of bias in at least one domain (95% CI 40.4 to 56.8%). Figure 3~~2~~ illustrates the rapid increase in number of QI RCTs published over time and the cumulative proportion of trials having at least one domain with high risk of bias up to a given year. In general, the line representing the proportion at high risk of bias runs parallel to the number of trials published, consistently accounting for almost half of the studies. Table 4 indicates a lack of significant trend over time in the proportion of trials with at least one domain with high risk of bias: these proportions were 46%, 44%, and 54% before 2002, between 2002 and 2006, and after 2006, respectively. Table 4 also demonstrates a lack of significant association between any of the study characteristics considered and presence of high risk of bias in at least one domain.

The sensitivity analysis, restricting studies defined as high risk of bias overall to those with high risk of bias in one of four domains (allocation sequence generation, allocation concealment, blinding, or selective outcome reporting) also revealed no trends over time – the proportions were 19%, 20%, and 20% before 2002, between 2002 and 2006 and after 2006, respectively (p=0.86).

~~Table 3 also demonstrates a lack of significant association between any of the study characteristics considered and presence of high risk of bias in at least one domain.~~

DISCUSSION

Main findings

Using the Cochrane EPOC Risk of Bias Tool,[6] we found that nearly half of RCTs focusing on diabetes had at least one domain at high risk of bias. The trials were most often at high risk of bias due to inadequate follow-up of participants, a lack of similarity at baseline across outcome measures or covariates, or inadequate blinding. We also noted that the majority of RCT reports failed to include an adequate description of the allocation process (i.e., sequence generation and allocation concealment were 'unclear'). To be interpreted appropriately, RCTs must be completely and transparently reported.[15,16] Our findings indicate that greater efforts are needed to ensure both adequate reporting and methodological conduct of diabetes QI trials.

We found that poor follow-up, baseline imbalances, and blinding were the most common sources of high risk of bias. Although these domains may be difficult fully control in QI trials, methodological approaches are available to mitigate and/or explore such causes of risk of bias. For example, sensitivity analyses may be used to explore the risk of bias related to loss to follow up, and risk of baseline imbalances in QI trials may be reduced through restricted randomization techniques, especially when trials are cluster-randomized with relatively few clusters. In addition, selective outcome reporting may be limited if more QI trial protocols were registered. Finally, although blinding may be particularly difficult to accomplish in QI trials, this should be clearly reported; if outcome assessment is not blinded, risk of bias could still be limited by using objective outcomes.

~~For example, although blinding may be particularly difficult to accomplish in QI trials, this should be clearly reported; risk of bias could still be limited by using objective outcomes.~~

Comparison to literature

A systematic review focusing on cluster r~~andomized~~andomised trials found minimal improvement over time in either reporting or methodological conduct.[17] We found no evidence for a difference in the proportion of cluster- r~~andomized~~andomised trials at high risk of bias compared to trials in which individuals were allocated. However, imbalance at baseline was a common source of potential bias in diabetes QI trials, possibly owing to inadequate use of restricted randomization in cluster trials.[18] Another systematic review included 35 studies covering a range of health-related fields assessing trends over time in quality criteria for RCTs.[5] Of these, 26 found improvement over time for at least one aspect of methodological quality. The domain most commonly noted to have improvement was allocation concealment, but the authors noted that this domain remained either poorly reported or inadequately performed in over half of the examined trials. We found a similarly low proportion of studies clearly reporting adequate allocation concealment, and no evidence of improvement over time.

Previous authors have noted that QI reports may not contain enough information to inform generalization and allow for replication in different clinical settings.[19] Standards for Quality Improvement Reporting (SQUIRE) guidelines suggest that investigators conducting trials use both SQUIRE and CONSORT to inform their manuscripts.[16] Journal editors should enforce the requirements of both SQUIRE and CONSORT for QI

RCTs, possibly by permitting detailed information to be posted as online appendices. Although it might seem onerous to force investigators to address all items in SQUIRE and CONSORT, the risks of poor reporting are substantial. Inadequate description of context could omit essential pre-conditions or important effect modifiers for a successful QI program, while incomplete description of the program itself might lead to failure due to partial implementation.

Strengths and limitations

To our knowledge, this is the largest analysis of risk of bias ever reported for health care QI RCTs and the only one to assess for trends over time. The findings are strengthened by the rigorous methods used to prepare the data for the systematic review.

QI evaluations have been criticized based on numerous criteria beyond the risk of bias domains, including short duration of intervention, lack of justification for intervention design, and poor generalizability.[2021] Some important components of methodological quality do not relate to bias (e.g. reporting of a sample size calculation). Thus, it is possible that studies at low risk of bias have important flaws with respect to methodology and/or reporting (and vice-versa), and it is possible that using other scales to assess study quality could have led to different results.[22] While the overall risk of bias assessment using the Cochrane Risk of Bias Tool has been shown to differentiate effect sizes (i.e., higher risk of bias studies usually have larger effect sizes),[10] studies at high risk of bias may still offer valuable knowledge for QI implementers. The merit of any given report will depend on the needs of the reader, while the current analysis provides an assessment of the progress in the literature as a whole.

Furthermore, we acknowledge that assigning trials with high risk of bias in a single domain a status of high risk of bias *overall* may be arguable. ~~For this reason, we assessed trends in individual domains in addition to the summary score and also conducted a *post-hoc* sensitivity analysis that applied an empirically based rule for assigning high risk of bias overall. Previous meta-analyses have found that high risk of bias in four specific domains, namely allocation sequence generation, allocation concealment, blinding, and selective outcome reporting are each associated with greater effect size.~~[22] ~~[24]O~~Nevertheless, our sensitivity analysis ~~considering studies with high risk of bias in any of these four (rather than all) domains to be at high risk of bias *overall*~~ led to the same conclusion: there has been no improvement over time in the proportion of trials at high risk of bias in this literature and no particular study characteristics were associated with high risk of bias.

Another potential limitation stems from our analytical approach regarding change over time; collapsing publication year into three timeframes (pre-2002, 2002-2006, 2007-2010) and testing for trends may have limited our power. These timeframes were chosen *a priori* based on the publication of important documents that we thought might affect the conduct and reporting of these trials. We felt the assumption of linear change over time underlying the Cochran-Armitage test for trend was appropriate and in keeping with our hypotheses (e.g. high and unclear risk of bias would decrease gradually over time, while low risk of bias would increase). Risk of type 2 error is tempered by the number of tests performed; the lack of a significant p-value for trend for any level of risk of bias in any domain supports our main conclusion. Finally, this review considered only RCTs from

Field Code Changed

the diabetes literature. It would have been preferable to evaluate a random sample of all QI trials, but adequate QI electronic literature searches have yet to be developed.[25]

Implications

Published trials testing QI in diabetes are frequently at high risk of bias, producing results that may not be replicable. Clinicians must scrutinize the internal validity of the results as a first step in the process of considering the application of clinical findings for particular patients. Our findings emphasize the need for policy-makers, managers, and/or clinical-administrators seeking to implement QI interventions to apply the same process.[3] It is likely that QI investigators publishing RCTs desire for their work to have a broad impact. To help them accomplish this, research funders and journal editors can play an important role by ensuring that QI trials are reported thoroughly and transparently and are designed in a manner that limits the potential for risk of bias.

CONFLICT OF INTEREST:

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The authors declare that: (i) NM, ACT, MT, IH, LT, DM, & JG received support from Ontario Ministry of Health and Long-term Care, and the Alberta Heritage Foundation for the original systematic review, but the funding agencies had no role in the study design, collection, analysis or interpretation of data, writing of the manuscript or in the decision to submit this manuscript for publication; (ii) NM, ACT, MT, IH, LT, DM, & JG have no relationships with any companies that might have an interest in the submitted work in the previous 3 years; (iii) their spouses, partners, or children have no financial relationships that may be relevant to the submitted work; and (iv) NM, ACT, MT, IH, LT, DM, & JG have no non-financial interests that may be relevant to the submitted work.

AUTHOR CONTRIBUTIONS

NI and ACT designed and coordinated the study, participated in data collection, data analysis, data interpretation, and drafted the manuscript. MT conducted the analysis and participated in data interpretation and drafting the manuscript. IH, LT, DM, and JMG helped to design the study and write the manuscript. All authors read and approved the final manuscript.

DATA SHARING

Data are available upon request from the corresponding authour.

SOURCES OF SUPPORT

NMI holds fellowship awards from the Canadian Institutes of Health Research (CIHR) and from the Department of Family and Community Medicine, University of Toronto.

ACT holds a CIHR/Drug Safety and Effectiveness Network new investigator award.

JMG and DM both hold Canada Research Chairs.

References

1. Tricco AC, Ivers NM, Grimshaw JM, et al. Effectiveness of quality improvement strategies on the management of diabetes: a systematic review and meta-analysis. *Lancet* 2012;379:2252-2261.

2. Pildal J, Hrobjartsson A, Jorgensen KJ, et al. Impact of allocation concealment on conclusions drawn from meta-analyses of ~~randomized~~randomised trials. *Int J Epidemiol* 2007;36:847-857.

3. Fan E, Laupacis A, Pronovost PJ, et al. How to use an article about quality improvement. *JAMA* 2010;304:2279-2287.

4. Higgins JP, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.

5. Falagas ME, Grigori T, Ioannidou E. A systematic review of trends in the methodological quality of ~~randomized~~randomised controlled trials in various research fields. *J Clin Epidemiol* 2009;62:227-231.

6. Cochrane Effective Practice and Organisation of Care Group. Suggested risk of bias criteria for EPOC reviews, 2012. http://epoc.cochrane.org/sites/epoc.cochrane.org/files/uploads/Suggested%20risk%20of%20bias%20criteria%20for%20EPOC%20reviews.pdf

7. SCImago. SJR — SCImago Journal & Country Rank, 2007.

8. Falagas ME, Kouranos VD, Arencibia-Jorge R, et al. Comparison of SCImago journal rank indicator with journal impact factor. *FASEB J* 2008;22:2623-2628.

9. Higgins JPT, Altman DG. Assessing risk of bias in included studies. In: Higgins JPT, Green S, ed. Cochrane handbook for systematic reviews of interventions: Wiley, 2008:187-241.

10. Hartling L, Ospina M, Liang Y, et al. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. *BMJ* 2009;339:b4012.

11. Ivers NM, Halperin IJ, Barnsley J, et al. Allocation techniques for balance at baseline in cluster r~~andomized~~andomised trials: a methodological review. *Trials* 2012;13:120.

12. Moher D, Schulz KF, Altman D, et al. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group r~~andomized~~andomised trials. *JAMA* 2001;285:1987-1991

13. Shojania KG, Ranji SR, McDonald KM, et al. Effects of quality improvement strategies for type 2 diabetes on glycemic control: a meta-regression analysis. *JAMA* 2006;296:427-440.

14. SAS 9.2 [program]. Cary, NC: SAS Institute Inc., 2008.

15. Simera I, Moher D, Hirst A, et al. Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR Network. *BMC Med* 2010;8:24.

16. Davidoff F, Batalden P, Stevens D, et al. Publication guidelines for quality improvement studies in health care: evolution of the SQUIRE project. *BMJ* 2009;338:a3152.

17. Ivers NM, Taljaard M, Dixon S, et al. Impact of CONSORT extension for cluster randomised trials on quality of reporting and study methodology: review of random sample of 300 trials, 2000-8. *BMJ* 2011;343:d5886.

18. Campbell MK, Elbourne DR, Altman DG, et al. CONSORT statement: extension to cluster randomised trials. *BMJ* 2004;328(7441):702-708.

19. Michie S, Fixsen D, Grimshaw JM, et al. Specifying and reporting complex behaviour change interventions: the need for a scientific method. *Implement Sci* 2009;4:40.

20. Alexander JA, Hearld LR. What can we learn from quality improvement research? A critical review of research methods. *Med Care Res Rev* 2009;66:235-271.

21. Eccles M, Grimshaw J, Walker A, et al. Changing the behavior of healthcare professionals: the use of theory in promoting the uptake of research findings. *J Clin Epidemiol* 2005;58:107-112.

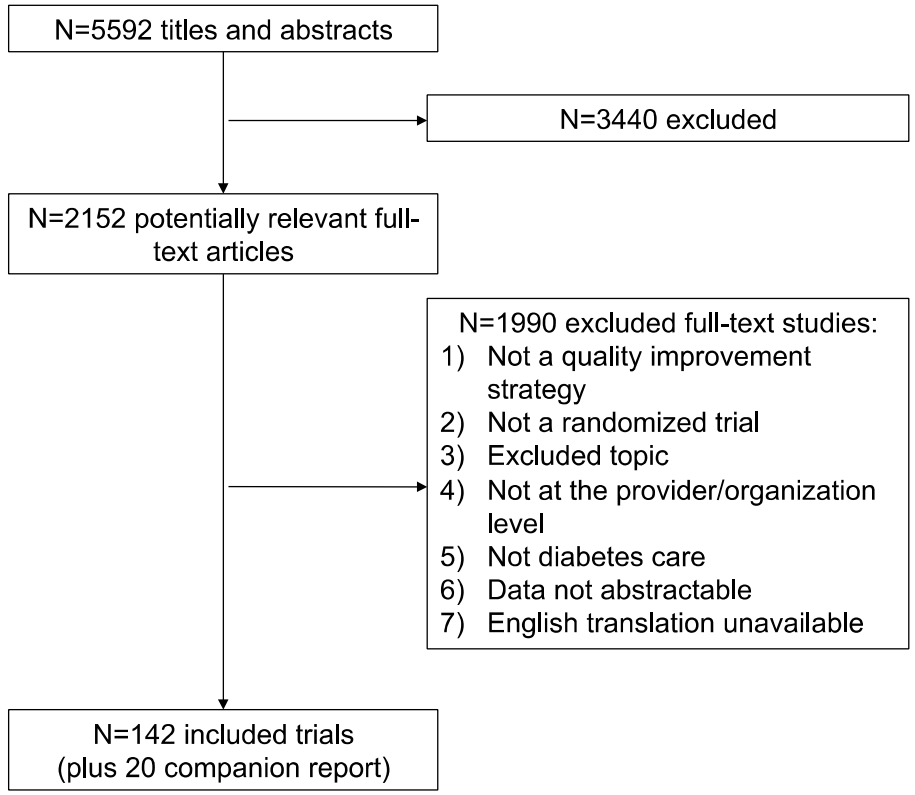
22. Armijo-Olivo S, Stiles CR, Hagen NA, et al. Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. *J Eval Clin Pract* 2012;18:12-18

23. Odgaard-Jensen J, Vist GE, Timmer A, et al. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database Syst Rev* 2011:MR000012.

24. Hempel S, Suttorp MJ, Miles JNV, et al. Empirical Evidence of Associations Between Trial Quality and Effect Size. Rockville (MD): Agency for Healthcare Research and Quality; 2011.

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

Figure 1: Study flow diagram



Formatted: Font: Bold

420

Figure 2~~1~~: percentage of studies judged to be at low, unclear, or high risk of bias in each risk of bias domain

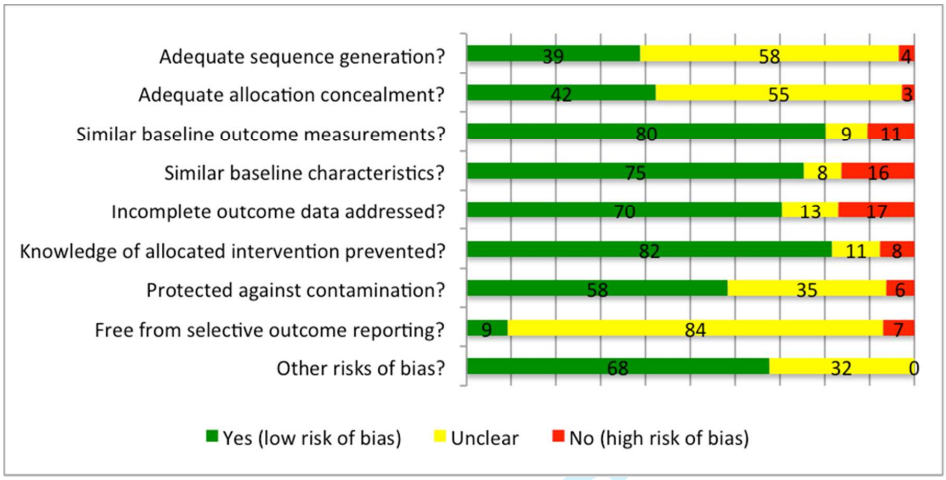


Figure 32: cumulative number of diabetes quality improvement trial publications at high risk of bias in any domain, 1990 to 2010.

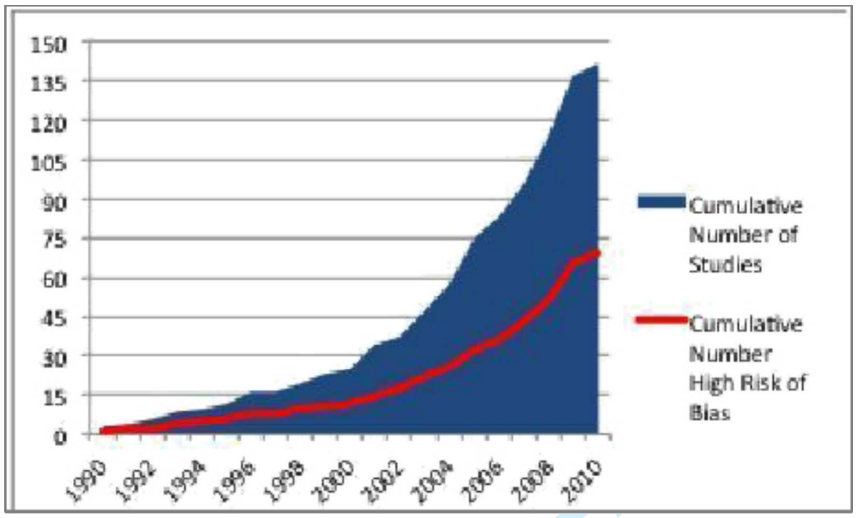


Table 1: Cochrane Effective Practice and organization of care (epoc) risk of bias assessment tool*

Risk of Bias Domain	Low Risk of Bias	High Risk of Bias	Unclear Risk of Bias
Was the allocation sequence adequately generated?	A random component in the sequence generation process is described (e.g. referring to a random number table)	Nonrandom method is used (e.g. performed by date of admission)	Not specified in the paper
Was the allocation adequately concealed?	The unit of allocation was by institution, team or professional and allocation was performed on all units at the start of the study; or if the unit of allocation was by patient or episode of care and there was some form of centralized randomization scheme, an on-site computer system or sealed opaque envelopes were used	Allocation was not adequately concealed	Not specified in the paper
Were baseline outcome measurements similar?	Performance or patient outcomes were measured prior to the intervention, and no important differences were present across study groups, or if imbalanced but appropriate adjusted analysis was performed	Important differences were present and not adjusted for in analysis	If no baseline measure of outcome
Were baseline characteristics similar?	Baseline characteristics of the study and control providers are reported and similar.	No report of characteristics in text or tables or if there are differences between control and intervention providers.	Not clear in the paper
Were incomplete outcome data adequately addressed?	Missing outcome measures were unlikely to bias the results	Missing outcome data was likely to bias the results.	Not specified in the paper
Was knowledge of the allocated interventions adequately prevented?	The authors state explicitly that the primary outcome variables were assessed blindly, or the outcomes are objective (e.g. length of hospital stay)	Outcomes were not assessed blindly and not objective.	Not specified in the paper.
Was the study free from selective outcome reporting?	There is no evidence that outcomes were selectively reported	Some important outcomes are omitted from the results	Not specified in the paper
Was the study adequately protected against contamination?	Allocation was by community, institution or practice and it is unlikely that the control group received the intervention.	It is likely that the control group received the intervention.	Communication between intervention and control professionals could have occurred
Was the study free from other risks of bias?	There is no evidence of other risk of biases		

* Adapted for ease of presentation. See http://epoc.cochrane.org/epoc-resources-review-authors for full explanation.

Table 2: Study and patient characteristics

Characteristic	Result
Patient RCTs, number (%)	94 (66.2)
Cluster RCTs, number (%)	48 (33.8)
Number of clusters, median [IQR]	29 [12, 57]
Numberof patients, median [IQR]	405.3 [203, 878]
Duration of intervention months, median [IQR]	12 [8.9, 15.0]
Mean age in years, median [IQR]	59.4 [54.9, 62.9]
Percent male, median [IQR]	49.8 [41.8, 55.9]
Type of diabetes N (%)	9 (6.3)
Type 1 diabetes	80 (56.3)
Type 2 diabetes	34 (23.9)
Type 1 and 2 diabetes	19 (13.4)
Type of diabetes unclear/NR	
Number of QIs per RCT median [IQR]	2 [0, 3.5]
Administrators of patient intervention(s) N (%)	
Primary care physician	30 (21.1)
Nurse	67 (47.2)
Pharmacist	19 (13.4)
Dietician	22 (15.5)
Psychiatrist	3 (2.1)
Psychologist	2 (1.4)
Ophthalmologist	2 (1.4)
Specialist/Endocrinologist	21 (14.8)
Other	49 (34.5)
Location of study N (%)	
United States	68 (47.9)
United Kingdom	14 (9.9)
Canada	11 (7.7)
Netherlands	8 (5.6)
South Korea	7 (4.9)
Australia	6 (4.2)
Denmark	3 (2.1)
Belgium	1 (0.7)
Israel	3 (2.1)
Spain	3 (2.1)
Norway	2 (1.4)
France	2 (1.4)
Germany	2 (1.4)
Italy	2 (1.4)
Switzerland	2 (1.4)
China	2 (1.4)
Ireland	1 (0.7)
New Zealand	1 (0.7)
Thailand	1 (0.7)
Taiwan	1 (0.7)
United Arab Emirates	1 (0.7)
Mexico	1 (0.7)

Notes: † All IQRs reported as the 25th and 75th percentiles, includes investigators and community workers.
Abbreviations: RCT r~~andomized~~andomised clinical trial, N number, IQR inter-quartile range, NA not applicable, NR not reported, QI quality improvement.

Formatted: Line spacing: single

Table 3: Trends over time in proportions of trials classified high, unclear, or low for each risk of bias domain

RISK OF BIAS DOMAIN	Pre-2002 N=37	2002-2006 N=46	2007-2010 N=59	P-value*	P-value^
Was the allocation sequence adequately generated?				0.41	0.43
Low	11 (30%)	19 (41%)	25 (42%)		
Unclear	24 (65%)	25 (55%)	33 (56%)		
High	2 (5%)	2 (4%)	1 (2%)		
Was the allocation adequately concealed?				1.00	0.82
Low	15 (40%)	20 (44%)	25 (42%)		
Unclear	21 (57%)	25 (54%)	32 (54%)		
High	1 (3%)	1 (2%)	2 (4%)		
Were baseline outcomes similar?				0.87	0.20
Low	31 (84%)	39 (85%)	44 (75%)		
Unclear	2 (5%)	3 (6%)	8 (13%)		
High	4 (11%)	4 (9%)	7 (12%)		
Were baseline characteristics similar?				0.16	0.57
Low	30 (81%)	34 (74%)	43 (73%)		
Unclear	3 (8%)	6 (13%)	3 (5%)		
High	4 (11%)	6 (13%)	13 (22%)		
Were incomplete outcome data adequately addressed?				0.17	0.70
Low	29 (78%)	33 (72%)	38 (64%)		
Unclear	3 (8%)	8 (17%)	7 (12%)		
High	5 (14%)	5 (11%)	14 (24%)		
Was knowledge of the allocated interventions prevented?				0.44	0.61
Low	32 (87%)	38 (83%)	46 (78%)		
Unclear	3 (8%)	5 (11%)	7 (12%)		
High	2 (5%)	3 (6%)	6 (10%)		
Was the study protected against contamination?				0.54	0.78
Low	25 (68%)	23 (50%)	35 (59%)		
Unclear	10 (27%)	21 (46%)	19 (32%)		
High	2 (5%)	2 (4%)	5 (9%)		
Was the study free from selective outcome reporting?				0.84	1.00
Low	3 (8%)	4 (9%)	6 (10%)		
Unclear	32 (87%)	37 (80%)	50 (85%)		
High	2 (5%)	5 (11%)	3 (5%)		
Was the study free from other risks of bias?				0.58	0.58
Low	27 (73%)	30 (65%)	39 (66%)		
Unclear	10 (27%)	16 (35%)	20 (34%)		
High	0	0	0		

* Exact Cochran-Armitage test for high versus low or unclear risk of bias in each domain except the last domain which was analyzed as low versus high or unclear due to absence of studies with high risk of bias.
^ Exact Cochran-Armitage test for reported (high or low risk of bias) or unreported (unclear risk of bias) in each domain.

Table 4: association between study characteristics and risk of bias

Characteristic	All studies, No.	Studies in high risk of bias in at least one domain No. (%)	P-value*
Year of publication			**0.37**
Pre-2002	37	17 (46%)	
2002-2006	46	20 (44%)	
2007-2010	59	32 (54%)	
Type of diabetes			**0.11**
Type 1	9	3 (33%)	
Type 2	80	36 (45%)	
Both	34	16 (47%)	
Unclear	19	14 (74%)	
Unit of Allocation			**0.24**
Patient	94	49 (52%)	
Cluster (e.g. provider/clinic)	48	20 (42%)	
Country/Setting			**0.62**
USA or Canada	79	41 (52%)	
UK or Western Europe	40	17 (43%)	
Other	23	11 (48%)	
Journal Impact Factor			**0.87**
Greater than 3 (median)	71	34 (47.9%)	
Less than 3 (median)	71	35 (49.3%)	
Effective Sample Size			**0.87**
Greater than 154 (median)	71	35 (49.3%)	
Less than 154 (median)	71	34 (47.9%)	
Intervention Type			**0.17**
Multifaceted (featuring more than one QI strategy)	124	63 (51%)	
Single intervention	18	6 (33%)	

* Comparing proportion of studies with at least one domain at high risk of bias against studies no domains at high risk of bias. For year of publication, Cochran-Armitage test for trend was conducted. For other study characteristics, chi-squared (or Fisher's exact) tests for categorical and Wilcoxon signed-rank tests for continuous variables were used

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	3-4
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	5
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	5-6
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	N/A
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	6-7
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	6, detailed strategy previously published
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	6-7
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	7
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	8-9
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	7-8
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	9-10

PRISMA 2009 Checklist

Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency ($e.g., I^2$) for each meta-analysis.	6-9

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	23-24
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	N/A
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	9-10 (flow chart previously published)
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Previously published
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	23-24
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Risk of bias data for each study available upon request.
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	N/A
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	23
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	N/A
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	10-12
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	13-14

PRISMA 2009 Checklist

Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	15
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	2

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: **www.prisma-statement.org**.

Page 2 of 2