



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

Journal:	<i>BMJ Open</i>
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
<b>Primary Subject Heading</b>:	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

SCHOLARONE™  
Manuscripts

1  
2  
3 **1 Quality improvement needed in quality improvement randomized trials:**

4  
5  
6 **2 Systematic review of interventions to improve care in diabetes**

7  
8 Noah M Ivers<sup>1</sup>, Andrea C Tricco<sup>2</sup>, Monica Taljaard<sup>3</sup>, Ilana Halperin<sup>4</sup>, Lucy Turner<sup>5</sup>,

9  
10 David Moher<sup>6</sup>, Jeremy M Grimshaw<sup>7</sup>

11  
12  
13  
14 <sup>1</sup>Department of Family and Community Medicine, Women's College Hospital-University  
15 of Toronto, 76 Grenville Street, Toronto, Ontario, M5S1B2, Canada, *Family physician*

16  
17  
18 <sup>2</sup>Li Ka Shing Knowledge Institute of St Michael's Hospital, 30 Bond Street, Toronto,  
19 Ontario, M5B 1W8, Canada, *Scientist*

20  
21  
22  
23 <sup>3</sup>Clinical Epidemiology Program, Ottawa Hospital Research Institute, 1053 Carling  
24 Avenue, Ottawa, Ontario, K1Y 4E9, Canada, *Scientist*

25  
26  
27  
28 <sup>4</sup>Division of Endocrinology and Metabolism, Department of Medicine, University of  
29 Toronto, 200 Elizabeth St, Toronto, Ontario, M5G2C4, Canada, *Endocrinology fellow*

30  
31  
32  
33 <sup>5</sup>Clinical Epidemiology Program, Ottawa Health Research Institute, Ottawa Hospital -  
34 General Campus, Centre for Practice-Changing Research 501 Smyth Road, Ottawa,  
35 Ontario, K1H 8L6, Canada, *Statistician*

36  
37  
38  
39  
40 <sup>6</sup>Clinical Epidemiology Program, Ottawa Health Research Institute, Ottawa Hospital -  
41 General Campus, Centre for Practice-Changing Research 501 Smyth Road, Ottawa,  
42 Ontario, K1H 8L6, Canada, *Scientist*

43  
44  
45  
46  
47 <sup>7</sup>Clinical Epidemiology Program, Ottawa Health Research Institute, Ottawa Hospital -  
48 General Campus, Centre for Practice-Changing Research 501 Smyth Road, Ottawa,  
49 Ontario, K1H 8L6, Canada, *Senior scientist*

1  
2  
3 22 **Correspondence to:** Noah Ivers, MD, Telephone: 4163236060, ext1 then 2, email:  
4  
5 23 noah.ivers@utoronto.ca, fax:4163236255.  
6  
7

8 24  
9

10  
11 25 **Ethical approval:** Not required.  
12

13  
14 26 **Data Sharing:** Statistical code and dataset are available from the corresponding author.  
15

16 27 **Word count:** 2532 (main text), 307 (abstract), 25 references, 4 tables, 2 figures.  
17

18 28 **Key words:** systematic review, quality improvement, risk of bias, diabetes mellitus,  
19  
20 29 randomized controlled trials  
21  
22

23  
24 30  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 31 **Article Summary**  
4

5 32 **Article focus**  
6

- 7  
8 33 • Reliable quality improvement research is needed to make decisions about initiating or  
9  
10 34 scaling up quality improvement strategies.  
11  
12 35 • The number of published quality improvement trials has increased rapidly over time.  
13  
14 36 • The quality of trials published in other areas of health seem to be improving over time  
15  
16 37 but the risk of bias in the quality improvement literature is uncertain  
17  
18  
19

20 38 **Key messages**  
21

- 22 39 • Nearly half of quality improvement trials for diabetes are at high risk of bias.  
23  
24 40 • The quality of quality improvement trials does not seem to be improving over time.  
25  
26 41 • Policy-makers, administrators, clinicians, and research funders must carefully  
27  
28 42 scrutinize the methods used in quality improvement trials to ensure evidence-based  
29  
30 43 quality improvement.  
31  
32  
33

34 44 **Strengths and limitations of this study**  
35

- 36  
37 45 • This is the largest systematic review of risk of bias in the quality improvement  
38  
39 46 literature and the only to assess for trends over time.  
40  
41 47 • The risk of bias tool does not capture all sources of methodological bias and poor  
42  
43 48 reporting interferes with the assessment of many domains.  
44  
45  
46 49 • The merits of any given trial report depends to some extent on the needs of the reader.  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 **50 Abstract**  
4

5  
6 51 Objective: Despite an increasing number of published trials of quality improvement (QI)  
7  
8 52 interventions in diabetes, little is known about the risk of bias in this literature.  
9

10  
11 53 Design: Secondary analysis of a systematic review.

12  
13 54 Data sources: Medline, the Cochrane Effective Practice and Organisation of Care (EPOC)  
14  
15 55 database (from inception to July 2010), and references of included studies.  
16

17  
18 56 Eligibility criteria: Randomized trials assessing 11 predefined QI strategies or financial  
19  
20 57 incentives targeting health systems, health-care professionals, or patients to improve  
21  
22 58 management of adult outpatients with diabetes.  
23

24  
25 59 Analysis: The risk of bias (low, unclear, or high) was assessed for the 142 trials in the  
26  
27 60 review across nine domains using the EPOC version of the Cochrane Risk of Bias Tool.

28  
29 61 We used Cochran-Armitage tests for trends to evaluate improvement over time

30  
31 62 Results: There was no significant improvement over time in any of the risk of bias  
32  
33 63 domains. Attrition bias (loss to follow up) was the most common source of bias, with 24  
34  
35 64 trials (17%) having high risk of bias due to incomplete outcome data. Inadequate  
36  
37 65 reporting frequently hampered risk of bias assessment: allocation sequence was unclear in  
38  
39 66 82 trials (58%) and allocation concealment was unclear in 78 trials (55%). Overall, 69  
40  
41 67 trials (49%) had at least one domain with high risk of bias. There were no significant  
42  
43 68 reductions in the proportions of studies that were unclear or at high risk of bias over time.  
44  
45

46  
47 69 Conclusion: Nearly half of the included QI trials in this review were judged to have high  
48  
49 70 risk of bias. Such trials have serious limitations that put the findings in question and  
50  
51 71 therefore inhibit evidence-based QI. There is a need to limit the potential for bias when  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 72 conducting QI trials and improve the quality of reporting of QI trials so that stakeholders  
4  
5 73 have adequate evidence for implementation.  
6  
7

## 8 74 **Introduction**

9  
10  
11 75 There is significant interest in quality improvement (QI) in health care, as evidenced by  
12  
13 76 the rapidly increasing number of randomized clinical trials (RCTs) of QI interventions,  
14  
15  
16 77 especially in the diabetes literature.<sup>1</sup> RCTs can provide a foundation for making  
17  
18 78 statements regarding causation, but the validity of trials varies widely; trials with  
19  
20  
21 79 adequate allocation concealment and blinding generally produce smaller effect sizes.<sup>2</sup>  
22  
23 80 Since internal validity in QI trials is a necessary precursor for application to other  
24  
25 81 settings,<sup>3</sup> the ‘risk of bias’ of the findings should be assessed to ascertain the utility of the  
26  
27  
28 82 trial results. When an RCT is deemed to have high risk of bias, the study’s findings  
29  
30  
31 83 become questionable.<sup>4</sup>  
32

33  
34 84 Evaluations to assess trends in methodological quality of RCTs have been conducted in  
35  
36 85 many fields of health care,<sup>5</sup> but no previous reviews have assessed risk of bias in QI  
37  
38 86 RCTs or whether risk of bias in QI RCTs has changed over time. Recently, we conducted  
39  
40  
41 87 a systematic review and meta-regression that included 142 RCTs evaluating QI strategies  
42  
43 88 to improve care for patients with diabetes.<sup>1</sup> In this secondary analysis of those data, we  
44  
45 89 aimed to examine the risk of bias of included studies using the Cochrane Risk of Bias  
46  
47  
48 90 tool developed by the Cochrane Effective Practice and Organisation of Care (EPOC)  
49  
50 91 group<sup>6</sup> and determine whether the proportion with high risk of bias decreased over time.  
51  
52 92 We also evaluated trial and publication characteristics that might be associated with high  
53  
54  
55 93 risk of bias.  
56  
57  
58  
59  
60

1  
2  
3 94 **METHODS**  
4  
5

6  
7 95 A detailed description of the methods used for searching, screening, and abstracting the  
8  
9 96 relevant data has been published<sup>1</sup> and is briefly summarized here.  
10

11  
12 97 *Search strategy*  
13

14  
15  
16 98 Studies were identified by searching MEDLINE and the Cochrane EPOC database (up to  
17  
18 99 July 2010), and screening references of included RCTs. The search strategy has been  
19  
20 100 previously published<sup>1</sup> and is available upon request.  
21  
22

23  
24 101 *Study selection*  
25

26  
27 102 RCTs examining one of eleven pre-defined QI strategies, and/or financial incentives,  
28  
29 103 targeting health systems and/or healthcare professionals for the management of adult  
30  
31 104 outpatients with diabetes were included. RCTs had to report at least one of the chosen  
32  
33 105 process of care measures (proportion of patients taking acetylsalicylic acid, statins, anti-  
34  
35 106 hypertensive medication, screened for retinopathy, screened for foot abnormalities,  
36  
37 107 monitored for renal function) or intermediate outcomes (glycosylated hemoglobin levels,  
38  
39 108 low-density lipoprotein cholesterol levels, diastolic and systolic blood pressure,  
40  
41 109 proportion of patients with controlled hypertension, proportion of patients who quit  
42  
43 110 smoking) for inclusion.  
44  
45  
46  
47  
48

49 111 *Data abstraction*  
50

51  
52  
53 112 A draft data abstraction form was developed and modified after a training exercise among  
54  
55 113 reviewers. Two reviewers abstracted relevant data for each RCT independently.  
56  
57  
58  
59  
60

1  
2  
3 114 Discrepancies were resolved by discussion or the involvement of a third reviewer.  
4  
5 115 Authors of the included RCTs were contacted to obtain further information for data items  
6  
7  
8 116 requiring clarification. Journal impact factors from journal citation reports (ISI Web of  
9  
10 117 Science, 2009) were obtained. When a journal's ranking was unavailable, we used the  
11  
12 118 impact ranking of the open access SMImago journal and country rank database, if  
13  
14 119 available.<sup>7</sup> This ranking is calculated using a similar formula and is strongly correlated  
15  
16 120 with the journal citation impact factor.<sup>8</sup>  
17  
18  
19

### 20 21 121 *Assessing Risk of Bias*

22  
23  
24 122 As the included trials tested QI interventions, the Cochrane EPOC Risk of Bias Tool<sup>6</sup> was  
25  
26 123 used to assess the risk of bias in each study. The standard Cochrane Risk of Bias Tool  
27  
28 124 includes an assessment of seven domains: sequence generation, allocation concealment,  
29  
30 125 blinding of participants and personnel, blinding of outcome assessment, incomplete  
31  
32 126 outcome data, selective reporting, and other. The Cochrane Handbook<sup>9</sup> provides  
33  
34 127 instructions for making judgments about the specific domains as high, unclear, or low  
35  
36 128 risk. When formulating summary assessments for each trial, classification of a study as  
37  
38 129 "high risk" indicates that bias could have affected the results, while unclear risk of bias  
39  
40 130 indicates that some doubt exists about the results, and low risk of bias indicates that bias  
41  
42 131 is unlikely to affect the results. It has been shown empirically that studies classified as  
43  
44 132 high risk using this tool are more likely to have larger effect sizes.<sup>10</sup>  
45  
46  
47  
48  
49

50  
51 133 The EPOC tool was adapted to account for the unique features of QI trials. (The  
52  
53 134 guidelines for applying the Cochrane EPOC tool are summarized in Table 1.) For  
54  
55 135 example, in many QI trials it is not possible to blind participants. In addition, QI trials



1  
2  
3 136 may require cluster-randomization to avoid contamination, but in cluster-randomized  
4  
5 137 trials balance at baseline is a particular concern.<sup>11</sup> Therefore, the EPOC tool uses the  
6  
7  
8 138 same approach as the general Cochrane Risk of Bias Tool, but requires an assessment of  
9  
10 139 bias in nine domains: sequence generation, allocation concealment, similarity of baseline  
11  
12 140 measurements, similarity of baseline characteristics, incomplete data, blinding of  
13  
14 141 outcome assessment, contamination, selective outcome reporting, and other. Assessment  
15  
16 142 was conducted independently by a SR methodologist (ACT) and a clinician (NMI) and  
17  
18 143 conflicts were resolved by discussion with an expert QI trialist (JMG).

## 144 ANALYSIS

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

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

1  
2  
3 158 for each RCT based on whether or not the study was classified as high risk of bias in at  
4  
5 159 least one domain. We tested for trend over time in the proportion at high risk of bias  
6  
7  
8 160 overall, hypothesizing that the proportion would decline over time. We used the same  
9  
10 161 year of publication categories and conducted Cochran-Armitage tests for trend of the  
11  
12 162 dichotomous indicator.

13  
14  
15  
16 163 In addition, we tested for associations between high risk of bias in at least one domain  
17  
18 164 and study characteristics chosen *a priori*: type of diabetes (type 1, type 2, both, unclear),  
19  
20 165 type of allocation (cluster randomized, patient randomized), country (USA or Canada,  
21  
22 166 UK or Western Europe, Other), type of intervention (single, multifaceted), journal impact  
23  
24 167 factor, effective sample size, and year of publication using Chi-squared tests (or Fisher's  
25  
26 168 exact tests, as appropriate) for categorical and Wilcoxon signed-rank tests for continuous  
27  
28 169 measures. We hypothesized that each of these characteristics may be associated with  
29  
30 170 studies at high risk of bias overall. All analyses were conducted in SAS Version 9.2.<sup>14</sup>  
31  
32  
33  
34  
35

## 36 171 **RESULTS**

37  
38  
39 172 We analyzed 142 studies, with 37 (26%) published before 2002, 46 (32%) between 2002  
40  
41 173 and 2006, and 59 (42%) between 2007 and 2010. These studies evaluated the effects of  
42  
43 174 QI interventions on 123,529 patients with diabetes. Trial and patient characteristics are  
44  
45 175 described in Table 2. The proportions of studies judged to be at low, unclear, or high risk  
46  
47 176 of bias for each domain are illustrated in Figure 1. The domains most commonly at high  
48  
49 177 risk of bias were outcome reporting bias (17%) and similarity across characteristics at  
50  
51 178 baseline (16%). A lack of similarity in outcome measures at baseline (10%), and lack of  
52  
53 179 adequate blinding (8%) were also relatively common domains with high risk of bias.  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 180 Studies were rarely at high risk of bias due to the allocation sequence generation (4%) or  
4  
5 181 allocation concealment (3%), but these domains were often unclearly reported (57% and  
6  
7 182 55% unclear, respectively). Selective outcome reporting was deemed unclear 84% of the  
8  
9 183 time because published protocols were rarely available and it was often plausible that  
10  
11 184 many more outcomes than those reported were measured. Table 3 indicates a lack of  
12  
13 185 significant trend over time in the proportion of trials at low versus unclear or high risk of  
14  
15 186 bias for any given domain.  
16  
17  
18  
19

20  
21 187 Overall, 48.6% (69/142) of the RCTs had a high risk of bias in at least one domain (95%  
22  
23 188 CI 40.4 to 56.8%). Figure 2 illustrates the rapid increase in number of QI RCTs published  
24  
25 189 over time and the cumulative proportion of trials having at least one domain with high  
26  
27 190 risk of bias up to a given year. In general, the line representing the proportion at high risk  
28  
29 191 of bias runs parallel to the number of trials published, consistently accounting for almost  
30  
31 192 half of the studies. Table 4 indicates a lack of significant trend over time in the proportion  
32  
33 193 of trials with at least one domain with high risk of bias: these proportions were 46%,  
34  
35 194 44%, and 54% before 2002, between 2002 and 2006, and after 2006, respectively. Table  
36  
37 195 3 also demonstrates a lack of significant association between any of the study  
38  
39 196 characteristics considered and presence of high risk of bias in at least one domain.  
40  
41  
42  
43  
44

## 45 197 **DISCUSSION**

### 46 198 *Main findings*

47  
48  
49 199 Using the Cochrane EPOC Risk of Bias Tool,<sup>6</sup> we found that nearly half of RCTs  
50  
51 200 focusing on diabetes had at least one domain at high risk of bias. The trials were most  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 201 often at high risk of bias due to inadequate follow-up of participants, a lack of similarity  
4  
5 202 at baseline across outcome measures or covariates, or inadequate blinding. We also noted  
6  
7  
8 203 that the majority of RCT reports failed to include an adequate description of the  
9  
10 204 allocation process (i.e., sequence generation and allocation concealment were ‘unclear’).  
11  
12 205 To be interpreted appropriately, RCTs must be completely and transparently reported.<sup>15 16</sup>  
13  
14  
15 206 Our findings indicate that greater efforts are needed to ensure both adequate reporting and  
16  
17 207 methodological conduct of diabetes QI trials. For example, although blinding may be  
18  
19 208 particularly difficult to accomplish in QI trials, this should be clearly reported; risk of  
20  
21 209 bias could still be limited by using objective outcomes.  
22  
23  
24

#### 25 210 *Comparison to literature*

26  
27  
28  
29 211 A systematic review focusing on cluster randomized trials found minimal improvement  
30  
31 212 over time in either reporting or methodological conduct.<sup>17</sup> We found no evidence for a  
32  
33 213 difference in the proportion of cluster randomized trials at high risk of bias compared to  
34  
35 214 trials in which individuals were allocated. However, imbalance at baseline was a common  
36  
37 215 source of potential bias in diabetes QI trials, possibly owing to inadequate use of  
38  
39 216 restricted randomization in cluster trials.<sup>18</sup> Another systematic review included 35 studies  
40  
41 217 covering a range of health-related fields assessing trends over time in quality criteria for  
42  
43 218 RCTs.<sup>5</sup> Of these, 26 found improvement over time for at least one aspect of  
44  
45 219 methodological quality. The domain most commonly noted to have improvement was  
46  
47 220 allocation concealment, but the authors noted that this domain remained either poorly  
48  
49 221 reported or inadequately performed in over half of the examined trials. We found a  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 222 similarly low proportion of studies clearly reporting adequate allocation concealment, and  
4  
5 223 no evidence of improvement over time.  
6  
7

8  
9 224 Previous authors have noted that QI reports may not contain enough information to  
10  
11 225 inform generalization and allow for replication in different clinical settings.<sup>19</sup> Standards  
12  
13 226 for Quality Improvement Reporting (SQUIRE) guidelines suggest that investigators  
14  
15 227 conducting trials use both SQUIRE and CONSORT to inform their manuscripts.<sup>16</sup>  
16  
17 228 Journal editors should enforce the requirements of both SQUIRE and CONSORT for QI  
18  
19 229 RCTs, possibly by permitting detailed information to be posted as online appendices.  
20  
21  
22

23  
24 230 *Strengths and limitations*  
25  
26

27  
28 231 To our knowledge, this is the largest analysis of risk of bias ever reported for health care  
29  
30 232 QI RCTs and the only one to assess for trends over time. The findings are strengthened  
31  
32 233 by the rigorous methods used to prepare the data for the systematic review. QI  
33  
34 234 evaluations have been criticized based on numerous criteria beyond the risk of bias  
35  
36 235 domains, including short duration of intervention, lack of justification for intervention  
37  
38 236 design, and poor generalizability.<sup>20 21</sup> Some important components of methodological  
39  
40 237 quality do not relate to bias (e.g. reporting of a sample size calculation). Thus, it is  
41  
42 238 possible that studies at low risk of bias have important flaws with respect to methodology  
43  
44 239 and/or reporting (and vice-versa), and it is possible that using other scales to assess study  
45  
46 240 quality could have led to different results.<sup>22</sup> While the overall risk of bias assessment  
47  
48 241 using the Cochrane Risk of Bias Tool has been shown to differentiate effect sizes (i.e.,  
49  
50 242 higher risk of bias studies usually have larger effect sizes),<sup>10</sup> studies at high risk of bias  
51  
52 243 may still offer valuable knowledge for QI implementers. The merit of any given report  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 244 will depend on the needs of the reader, while the current analysis provides an assessment  
4  
5 245 of the progress in the literature as a whole. Furthermore, we acknowledge that assigning  
6  
7  
8 246 trials with high risk of bias in a single domain a status of high risk of bias *overall* may be  
9  
10 247 arguable. For this reason, we assessed trends in individual domains in addition to the  
11  
12 248 summary score and also conducted a *post-hoc* sensitivity analysis that applied an  
13  
14 249 empirically based rule for assigning high risk of bias overall. Previous meta-analyses  
15  
16  
17 250 have found that high risk of bias in four specific domains, namely allocation sequence  
18  
19 251 generation, allocation concealment, blinding, and selective outcome reporting are each  
20  
21 252 associated with greater effect size.<sup>22-24</sup> Our sensitivity analysis considering studies with  
22  
23 253 high risk of bias in any of these four (rather than all) domains to be at high risk of bias  
24  
25  
26 254 *overall* led to the same conclusion: there has been no improvement over time in the  
27  
28 255 proportion of trials at high risk of bias in this literature and no particular study  
29  
30 256 characteristics were associated with high risk of bias. Another potential limitation stems  
31  
32 257 from our analytical approach regarding change over time; collapsing publication year into  
33  
34 258 three timeframes (pre-2002, 2002-2006, 2007-2010) and testing for trends may have  
35  
36 259 limited our power. These timeframes were chosen *a priori* based on the publication of  
37  
38 260 important documents that we thought might affect the conduct and reporting of these  
39  
40 261 trials. We felt the assumption of linear change over time underlying the Cochran-  
41  
42 262 Armitage test for trend was appropriate and in keeping with our hypotheses (e.g. high and  
43  
44 263 unclear risk of bias would decrease gradually over time, while low risk of bias would  
45  
46 264 increase). Risk of type 2 error is tempered by the number of tests performed; the lack of a  
47  
48 265 significant p-value for trend for any level of risk of bias in any domain supports our main  
49  
50 266 conclusion. Finally, this review considered only RCTs from the diabetes literature. It  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 267 would have been preferable to evaluate a random sample of all QI trials, but adequate QI  
4  
5 268 electronic literature searches have yet to be developed.<sup>25</sup>  
6  
7

8  
9 269 *Implications*  
10

11  
12 270 Published trials testing QI in diabetes are frequently at high risk of bias, producing results  
13  
14 271 that may not be replicable. Clinicians must scrutinize the internal validity of the results as  
15  
16 272 a first step in the process of considering the application of clinical findings for particular  
17  
18 273 patients. Our findings emphasize the need for policy-makers, managers, and/or clinical-  
19  
20 274 administrators seeking to implement QI interventions to apply the same process.<sup>3</sup> It is  
21  
22 275 likely that QI investigators publishing RCTs desire for their work to have a broad impact.  
23  
24 276 To help them accomplish this, research funders and journal editors can play an important  
25  
26 277 role by ensuring that QI trials are reported thoroughly and transparently and are designed  
27  
28 278 in a manner that limits the potential for risk of bias.  
29  
30  
31  
32  
33

34 279  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 280 **CONFLICT OF INTEREST:**  
4

5 281 This research received no specific grant from any funding agency in the public,  
6  
7  
8 282 commercial, or not-for-profit sectors. The authors declare that: (i) NM, ACT, MT, IH,  
9  
10 283 LT, DM, & JG received support from Ontario Ministry of Health and Long-term Care,  
11  
12 284 and the Alberta Heritage Foundation for the original systematic review, but the funding  
13  
14 285 agencies had no role in the study design, collection, analysis or interpretation of data,  
15  
16 286 writing of the manuscript or in the decision to submit this manuscript for publication; (ii)  
17  
18 287 NM, ACT, MT, IH, LT, DM, & JG have no relationships with any companies that might  
19  
20 288 have an interest in the submitted work in the previous 3 years; (iii) their spouses,  
21  
22 289 partners, or children have no financial relationships that may be relevant to the submitted  
23  
24 290 work; and (iv) NM, ACT, MT, IH, LT, DM, & JG have no non-financial interests that  
25  
26 291 may be relevant to the submitted work.  
27  
28  
29  
30

31 292 **AUTHOR CONTRIBUTIONS**  
32

33  
34 293 NI and ACT designed and coordinated the study, participated in data collection, data  
35  
36 294 analysis, data interpretation, and drafted the manuscript. MT conducted the analysis and  
37  
38 295 participated in data interpretation and drafting the manuscript. IH, LT, DM, and JMG  
39  
40 296 helped to design the study and write the manuscript. All authors read and approved the  
41  
42 297 final manuscript.  
43  
44

45 298 **DATA SHARING**  
46

47  
48 299 Data are available upon request from the corresponding author.  
49

50 300 **SOURCES OF SUPPORT**  
51

52  
53 301 NMI holds fellowship awards from the Canadian Institutes of Health Research (CIHR)  
54  
55 302 and from the Department of Family and Community Medicine, University of Toronto.  
56  
57  
58  
59  
60



- 1  
2  
3 303 ACT holds a CIHR/Drug Safety and Effectiveness Network new investigator award.  
4  
5  
6 304 JMG and DM both hold Canada Research Chairs.  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

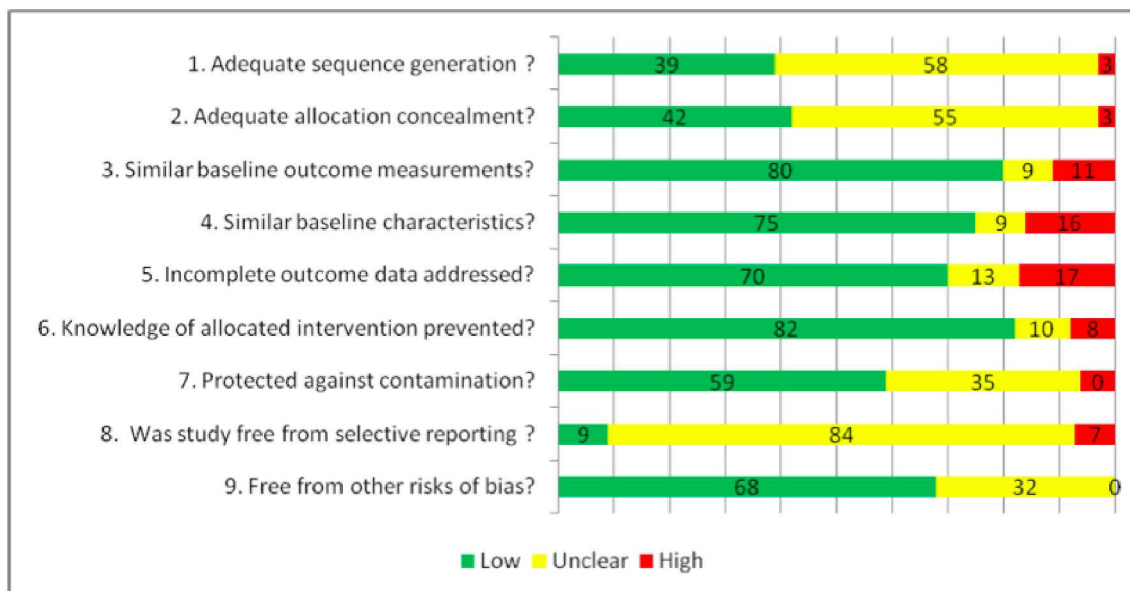
## References

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.

- 1  
2  
3 326 9. Higgins JPT, Altman DG. Assessing risk of bias in included studies. In: Higgins JPT,  
4  
5 327 Green S, ed. *Cochrane handbook for systematic reviews of interventions*: Wiley,  
6  
7 328 2008:187-241.  
8  
9  
10 329 10. Hartling L, Ospina M, Liang Y, et al. Risk of bias versus quality assessment of  
11  
12 330 randomised controlled trials: cross sectional study. *BMJ* 2009;339:b4012.  
13  
14  
15 331 11. Ivers NM, Halperin IJ, Barnsley J, et al. Allocation techniques for balance at baseline  
16  
17 332 in cluster randomized trials: a methodological review. *Trials* 2012;13:120.  
18  
19  
20 333 12. Moher D, Schulz KF, Altman D, et al. The CONSORT statement: revised  
21  
22 334 recommendations for improving the quality of reports of parallel-group randomized  
23  
24 335 trials. *JAMA* 2001;285:1987-1991  
25  
26  
27 336 13. Shojania KG, Ranji SR, McDonald KM, et al. Effects of quality improvement  
28  
29 337 strategies for type 2 diabetes on glycemic control: a meta-regression analysis. *JAMA*  
30  
31 338 2006;296:427-440.  
32  
33  
34 339 14. SAS 9.2 [program]. Cary, NC: SAS Institute Inc., 2008.  
35  
36 340 15. Simera I, Moher D, Hirst A, et al. Transparent and accurate reporting increases  
37  
38 341 reliability, utility, and impact of your research: reporting guidelines and the EQUATOR  
39  
40 342 Network. *BMC Med* 2010;8:24.  
41  
42  
43 343 16. Davidoff F, Batalden P, Stevens D, et al. Publication guidelines for quality  
44  
45 344 improvement studies in health care: evolution of the SQUIRE project. *BMJ*  
46  
47 345 2009;338:a3152.  
48  
49  
50 346 17. Ivers NM, Taljaard M, Dixon S, et al. Impact of CONSORT extension for cluster  
51  
52 347 randomised trials on quality of reporting and study methodology: review of random  
53  
54 348 sample of 300 trials, 2000-8. *BMJ* 2011;343:d5886.  
55  
56  
57  
58  
59  
60

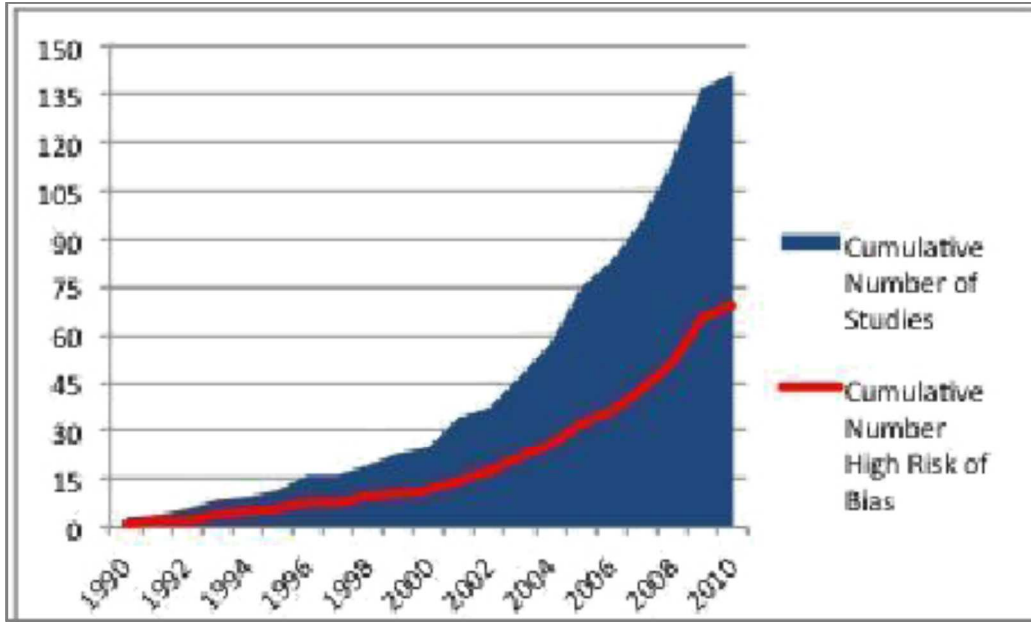
- 1  
2  
3 349 18. Campbell MK, Elbourne DR, Altman DG, et al. CONSORT statement: extension to  
4  
5 350 cluster randomised trials. *BMJ* 2004;328(7441):702-708.  
6  
7  
8 351 19. Michie S, Fixsen D, Grimshaw JM, et al. Specifying and reporting complex  
9  
10 352 behaviour change interventions: the need for a scientific method. *Implement Sci*  
11  
12 353 2009;4:40.  
13  
14  
15 354 20. Alexander JA, Hearld LR. What can we learn from quality improvement research? A  
16  
17 355 critical review of research methods. *Med Care Res Rev* 2009;66:235-271.  
18  
19  
20 356 21. Eccles M, Grimshaw J, Walker A, et al. Changing the behavior of healthcare  
21  
22 357 professionals: the use of theory in promoting the uptake of research findings. *J Clin*  
23  
24 358 *Epidemiol* 2005;58:107-112.  
25  
26  
27 359 22. Armijo-Olivo S, Stiles CR, Hagen NA, et al. Assessment of study quality for  
28  
29 360 systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and  
30  
31 361 the Effective Public Health Practice Project Quality Assessment Tool: methodological  
32  
33 362 research. *J Eval Clin Pract* 2012;18:12-18  
34  
35  
36 363 23. Odgaard-Jensen J, Vist GE, Timmer A, et al. Randomisation to protect against  
37  
38 364 selection bias in healthcare trials. *Cochrane Database Syst Rev* 2011:MR000012.  
39  
40  
41 365 24. Hempel S, Suttorp MJ, Miles JNV, et al. Empirical Evidence of Associations  
42  
43 366 Between Trial Quality and Effect Size. Rockville (MD): Agency for Healthcare Research  
44  
45 367 and Quality; 2011.  
46  
47  
48 368 25. Hempel S, Rubenstein LV, Shanman RM, et al. Identifying quality improvement  
49  
50 369 intervention publications--a comparison of electronic search strategies. *Implement Sci*  
51  
52 370 2011;6:85.  
53  
54  
55  
56  
57  
58  
59  
60

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



review only

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



Review only

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

<b>Risk of Bias Domain</b>	<b>Low Risk of Bias</b>	<b>High Risk of Bias</b>	<b>Unclear Risk of Bias</b>
<b>Was the allocation sequence adequately generated?</b>	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
<b>Was the allocation adequately concealed?</b>	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
<b>Were baseline outcome measurements similar?</b>	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
<b>Were baseline characteristics similar?</b>	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
<b>Were incomplete outcome data adequately addressed?</b>	Missing outcome measures were unlikely to bias the results	Missing outcome data was likely to bias the results.	Not specified in the paper
<b>Was knowledge of the allocated interventions adequately prevented?</b>	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.
<b>Was the study free from selective outcome reporting?</b>	There is no evidence that outcomes were selectively reported	Some important outcomes are omitted from the results	Not specified in the paper
<b>Was the study adequately protected against contamination?</b>	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
<b>Was the study free from other risks of bias?</b>	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 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*
<b>Was the allocation sequence adequately generated?</b>				<b>0.24</b>
Low	11 (30%)	19 (41%)	25 (42%)	
Unclear	24 (65%)	25 (55%)	33 (56%)	
High	2 (5%)	2 (4%)	1 (2%)	
<b>Was the allocation adequately concealed?</b>				<b>0.88</b>
Low	15 (40%)	20 (44%)	25 (42%)	
Unclear	21 (57%)	25 (54%)	32 (54%)	
High	1 (3%)	1 (2%)	2 (4%)	
<b>Were baseline outcomes similar?</b>				<b>0.22</b>
Low	31 (84%)	39 (85%)	44 (75%)	
Unclear	2 (5%)	3 (6%)	8 (13%)	
High	4 (11%)	4 (9%)	7 (12%)	
<b>Were baseline characteristics similar?</b>				<b>0.39</b>
Low	30 (81%)	34 (74%)	43 (73%)	
Unclear	3 (8%)	6 (13%)	3 (5%)	
High	4 (11%)	6 (13%)	13 (22%)	
<b>Were incomplete outcome data adequately addressed?</b>				<b>0.14</b>
Low	29 (78%)	33 (72%)	38 (64%)	
Unclear	3 (8%)	8 (17%)	7 (12%)	
High	5 (14%)	5 (11%)	14 (24%)	
<b>Was knowledge of the allocated interventions prevented?</b>				<b>0.29</b>
Low	32 (87%)	38 (83%)	46 (78%)	
Unclear	3 (8%)	5 (11%)	7 (12%)	
High	2 (5%)	3 (6%)	6 (10%)	
<b>Was the study protected against contamination?</b>				<b>0.55</b>
Low	25 (68%)	23 (50%)	35 (59%)	
Unclear	10 (27%)	21 (46%)	19 (32%)	
High	2 (5%)	2 (4%)	5 (9%)	
<b>Was the study free from selective outcome reporting?</b>				<b>0.72</b>
Low	3 (8%)	4 (9%)	6 (10%)	
Unclear	32 (87%)	37 (80%)	50 (85%)	
High	2 (5%)	5 (11%)	3 (5%)	
<b>Was the study free from other risks of bias?</b>				<b>0.52</b>
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*
<b>Year of publication</b>			<b>0.37</b>
Pre-2002	37	17 (46%)	
2002-2006	46	20 (44%)	
2007-2010	59	32 (54%)	
<b>Type of diabetes</b>			<b>0.11</b>
Type 1	9	3 (33%)	
Type 2	80	36 (45%)	
Both	34	16 (47%)	
Unclear	19	14 (74%)	
<b>Unit of Allocation</b>			<b>0.24</b>
Patient	94	49 (52%)	
Cluster (e.g. provider/clinic)	48	20 (42%)	
<b>Country/Setting</b>			<b>0.62</b>
USA or Canada	79	41 (52%)	
UK or Western Europe	40	17 (43%)	
Other	23	11 (48%)	
<b>Journal Impact Factor</b>			<b>0.87</b>
Greater than 3 (median)	71	34 (47.9%)	
Less than 3 (median)	71	35 (49.3%)	
<b>Effective Sample Size</b>			<b>0.87</b>
Greater than 154 (median)	71	35 (49.3%)	
Less than 154 (median)	71	34 (47.9%)	
<b>Intervention Type</b>			<b>0.17</b>
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 #
<b>TITLE</b>			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
<b>ABSTRACT</b>			
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
<b>INTRODUCTION</b>			
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
<b>METHODS</b>			
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
----------------------	----	---	-----

Page 1 of 2

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
<b>FUNDING</b>			
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](http://www.prisma-statement.org).

Page 2 of 2

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

1  
2  
3 RCTS. How was the use of CONSORT and/or SQUIRE associated with bias and QI success?  
4 4. One contributing factor is the role of extramural funding versus unfunded, small locally run QI  
5 interventions with "randomization". Please provide data on the QI trials as to whether or not they  
6 had extramural funding: e.g., did funding play a role in the design (and thus bias) and outcome of  
7 the study?

8 *These suggestions reflect a desire for further data abstraction from 142 published trials; this*  
9 *would be a daunting task even if the variables sought were well reported. However, we do not*  
10 *believe that the answers to the questions posed by this reviewer will be easily answered, as the*  
11 *variables are not likely to be readily accessible in many manuscripts. Improved reporting may*  
12 *reduce the number of domains marked 'unclear' but would not be expected to reduce the number*  
13 *of trials marked as high risk of bias. Part of the reason to encourage adherence to reporting*  
14 *guidelines would be to ensure that data from future trials may be more consistently extracted for*  
15 *methodological reviews in order to continue to monitor the state of the field. We have added a*  
16 *comment accordingly in the discussion.*

17  
18  
19 5. Expand the discussion section to make concrete recommendations on QI trial design,  
20 evaluation, and reporting. You begin this with the SQUIRE/CONSORT statements, but could be  
21 expanded to make a statement about where the field needs to go.

22 *The discussion section has been expanded.*

### 23 24 25 **Reviewer 2 –**

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

38  
39  
40 Study characteristics are reported in Table 1. This is not very detailed and further details on the  
41 included studies have to be assessed in the companion paper published in the Lancet.

42 *We have added to the Table a description of the QI strategies used.*

43  
44  
45 “journal editors should enforce the requirements of both SQUIRE and CONSORT for QI RCTs,  
46 possibly by permitting detailed information to be posted as online appendices” should be  
47 discussed a little further, as this introduces a different issue (contextualisation, external validity)  
48 and clearly raises some questions regarding feasibility.

49 *We have expanded the discussion section with respect to our recommendations.*

### 50 51 **Reviewer 3 –**

52 The full list of studies included in the analysis would have been useful. I reviewed reference 1 to  
53 gain a better understanding of the studies reviewed but was unable to access the online  
54 supplementary files.

55 *The supplementary files are available via Lancet online. We could seek permission to include the*  
56 *list of studies included as a supplementary file for this manuscript if the editors deem it*  
57 *necessary.*

1  
2  
3  
4 A table listing the QI strategies would be useful.  
5 *This has been done.*  
6

7  
8 Selective outcome reporting was unclear in 85% of studies. I question the relevance of this  
9 measure of bias in a QI study.

10 *We followed the rules set forth by the Cochrane Effective Practice and Organization of Care*  
11 *group in conducting this extraction. We attempted to emphasize that 'unclear' does not mean*  
12 *'high' risk of bias and for this reason we did some analysis examining trends independently for*  
13 *'low', 'unclear', or 'high' risk of bias.*

14 *We believe that this finding emphasizes the unique nature of QI trials, where there are often many*  
15 *outcomes measured since the interventions and goals are often multifaceted. It is conceivable that*  
16 *investigators might only report a selection of outcomes measured and that these may be more*  
17 *likely to be positive outcomes. For instance, consider the fact that more trials reported systolic*  
18 *BP than diastolic BP, when it seems inconceivable that they did not measure both. However, most*  
19 *studies do not publish protocols (nor do they reliably register trials in a complete manner)*  
20 *making it unclear whether some outcomes measured were not reported.*  
21

22  
23 **Question from the Editors:**

24 We were rather surprised that this analysis was not included in the original Lancet submission.  
25 Would you be willing to tell me why that was?

26 *The analyses in this manuscript reflect a desire to understand progression of the literature with*  
27 *respect to the methodological conduct of trials over time, while the Lancet review was concerned*  
28 *with the effect size of diabetes QI interventions. Certainly an analysis of risk of bias is standard*  
29 *in SRs, but an analysis of time trends in methodology/reporting is not. We believe our findings in*  
30 *this manuscript suggest a need to carefully (re-)consider the state of the science of quality*  
31 *improvement as a whole given the preponderance of substandard trials in the literature. As*  
32 *quality improvement trials that involve clinicians and their patients continue to increase in*  
33 *prevalence, we continue to believe, as mentioned by one of the referees, that the issue is relevant*  
34 *to a broad range of policy-makers, administrators, investigators, and clinicians.*  
35

36  
37 Thank you again for considering this article at BMJ Open.  
38

39 Given the ever-increasing resources dedicated to conducting such trials and the increasing interest  
40 amongst health care decision makers and stakeholders in the results of such trials, we believe the  
41 findings raise important concerns that require broad dissemination.  
42

43 Sincerely,  
44

45  
46  
47  
48  
49  
50  
51



51 Noah M Ivers MD CCFP PhD(c)

52 on behalf of the study co-authors: Andrea C Tricco, Monica Taljaard, Ilana Halperin, Lucy  
53 Turner, David Moher, and Jeremy M Grimshaw  
54



1  
2  
3 Paper: BMJ.2012.008810  
4

5 Decision: rejection  
6

7 Detailed comments from the meeting:  
8

9  
10 The committee echoed that of two out of three of the reviewers who felt the work is more  
11 valuable to a specialised reader with an interest in methodology, rather than the BMJ. We were  
12 rather surprised that this analysis was not included in the original Lancet submission. Would you  
13 be willing to tell me why that was?  
14

15 Please view the comments of the independent reviewers which are included at the end of this  
16 email.  
17

18 You will see that one of the reviewers was more positive than we were about the paper's  
19 suitability for the BMJ.  
20

21 Our main problem with the paper was that we did not think it added enough, for general readers,  
22 to what is already known about ....  
23

24 We did not have any specific criticism of the design or methods.  
25  
26  
27

28 Reviewer Comments:  
29

30 Reviewer: 1  
31

32 Recommendation:  
33  
34

35 Comments:

36 Dr. Ivers and colleagues have compiled a well-written evaluation of the quality improvement  
37 literature for diabetes by systematically reviewing 142 QI trials and providing a thorough analysis  
38 of on the role of bias. Most concerning, the author demonstrate that QI trials with a high degree  
39 of bias are rapidly increasing over the past 20 years whereby approximately 50% of the published  
40 trials today have a high degree of bias. This reviewer would encourage the authors to investigate  
41 the following improvements to their analysis:  
42

43 1. You have conducted excellent work in characterizing bias in these QI trials, however can you  
44 tie the high degree of bias to QI success?  
45

46 2. You have very rich and robust data abstracted from these trials. Could you use the data  
47 elements from Table 2 to conduct a meta-regression on QI success?  
48

49 3. You rightly conclude CONSORT and SQUIRE guidelines should be used for forthcoming QI  
50 RCTs; please provide data on the use of CONSORT and SQUIRE in the published QI  
51 RCTs. How was the use of CONSORT and/or SQUIRE associated with bias and QI success?  
52  
53

54 4. One contributing factor is the role of extramural funding versus unfunded, small locally run QI  
55 interventions with "randomization". Please provide data on the QI trials as to whether or not they  
56 had extramural funding: e.g., did funding play a role in the design (and thus bias) and outcome of  
57 the study?  
58  
59  
60

1  
2  
3  
4  
5 5. Expand the discussion section to make concrete recommendations on QI trial design,  
6 evaluation, and reporting. You begin this with the SQUIRE/CONSORT statements, but could be  
7 expanded to make a statement about where the field needs to go.

8  
9 All in all, a excellent analysis of the QI literature and of strong interest to the BMJ readership  
10 across the spectrum from provider, administrator to researcher and editors.

11  
12 Thank you

13  
14  
15 Additional Questions:

16 Please enter your name: Jeremiah Brown

17  
18 Job Title: Assistant Professor of Health Policy and Clinical Practice

19  
20 Institution: The Dartmouth Institute for Health Policy and Clinical Practice

21  
22 Reimbursement for attending a symposium?: No

23  
24 A fee for speaking?: No

25  
26 A fee for organising education?: No

27  
28 Funds for research?: Yes

29  
30 Funds for a member of staff?: No

31  
32 Fees for consulting?: No

33  
34  
35 Have you in the past five years been employed by an organisation that may  
36 in any way gain or lose financially from the publication of this paper?: No

37  
38 Do you hold any stocks or shares in an organisation that may in any way  
39 gain or lose financially from the publication of this paper?: No

40  
41 If you have any competing interests (either as indicated above or any other financial or non-  
42 financial interests) please declare them here: Agency for Healthcare Research and Quality  
43 (AHRQ) grant support for QI research  
44  
45

46  
47  
48 Reviewer: 2

49  
50 Recommendation:

51  
52 Comments:

53 Dear Editor and Authors,

54  
55 Thank you for the opportunity to review this interesting manuscript. Please see my comments  
56 below.  
57  
58  
59  
60

1  
2  
3 Sincerely Yours,  
4

5 Oliver Groene, PhD MSc MA, Lecturer in Health Services Research, London School of Hygiene  
6 and Tropical Medicine  
7  
8

9  
10  
11 The manuscript “Quality improvement needed in quality improvement randomized trials:  
12 Systematic review of interventions to improve care in diabetes” addresses the important issue of  
13 risk of bias in the literature. This is of relevance as the literature on quality improvement  
14 interventions is increasing and a risk-of-bias assessment, according to the authors, hasn’t been  
15 conducted in the field of QI trials. The authors further assess whether there is a time- trend in risk  
16 of bias, e.g. that more recent research publications are less subject to bias than earlier pieces of  
17 research. This review was prepared linked to another systematic review recently published by the  
18 author team (“Tricco AC et al. Effectiveness of quality improvement strategies on the  
19 management of diabetes: a systematic review and meta-analysis. Lancet 2012, 379: 2252-2261”).  
20

21  
22 In principle, I have little to criticize: this review is clearly focused, well conducted and written  
23 concisely, and uses frameworks and tools that are widely accepted (e.g EPOC QI strategies,  
24 Cochrane-EPOC Risk of Bias tool). It is reasonable to assume a time-trend in reported bias, as  
25 journals have adopted (to varying degrees) CONSORT and similar statements over the last couple  
26 of years. However, I am not so convinced of the time trend analysis as the groupings of  
27 publication year lead to loss of information and seem to suggest a significant impact of the  
28 publication of CONSORT and a previous systematic review in 2006. Unfortunately, rarely a  
29 single publication has such an impact.  
30

31  
32 This study reports that “nearly half of the included QI trials” were judged to have a high risk of  
33 bias. This bias has not improved over time.  
34

35 \* Originality - does the work add enough to what is already in the published literature? If so, what  
36 does it add? If not, please cite relevant references.  
37

38  
39 The study is well conducted and adds nuances to the existing literature. Risk of bias in RCTs is a  
40 well-known fact and the authors demonstrate that this also applies to QI interventions. This in  
41 itself, I would argue, is an important message, but probably more relevant to the audience in the  
42 quality improvement field, rather than in general medicine. The authors claim that this is the only  
43 analysis that assesses trend of bias over time (page 12, line 225). If this is the case, then it would  
44 be worthwhile to pursue such an analysis on a broader literature base, not only QI trials in the  
45 field of diabetes care.  
46

47 \* Importance of work to general readers - does this work matter to clinicians, patients, teachers,  
48 or policymakers? Is a general journal the right place for it?  
49

50  
51 See comments above: the study has its merits, but I would think that a more specialised journal in  
52 the field of quality improvement would be more appropriate. Risk of bias in RCTs is well-known  
53 and the time trend analysis reported here is limited.  
54

55 \* Scientific reliability

56 Research Question - clearly defined and appropriately answered?

57 Yes, clear, sound and appropriately answered, except the general concern regarding the  
58 assessment of time trend.  
59  
60

1  
2  
3  
4 \* Overall design of study - adequate?

5 Yes, except limitation in time-trend analysis  
6  
7

8 \* Participants studied - adequately described and their conditions defined?

9 Study characteristics are reported in Table 1. This is not very detailed and further details on the  
10 included studies have to be assessed in the companion paper published in the Lancet.  
11

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

14 Yes  
15

16 \* Results - answer the research question? Credible? Well presented?

17 Yes  
18

19 \* Interpretation and conclusions - warranted by and sufficiently derived from/focused on the  
20 data?

21 Message clear?  
22

23  
24 Yes, and the authors clearly discuss the key limitations of the paper. The conclusion that “journal  
25 editors should enforce the requirements of both SQUIRE and CONSORT for QI RCTs, possibly  
26 by permitting detailed information to be posted as online appendices” should be discussed a little  
27 further, as this introduces a different issue (contextualisation, external validity) and clearly raises  
28 some questions regarding feasibility.  
29

30 \* References - up to date and relevant? Any glaring omissions?

31 No  
32

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

34 yes  
35  
36  
37

38 Additional Questions:

39 Please enter your name: Oliver Groene  
40

41 Job Title: 1. Lecturer Health Services Research, 2. Senior methodologist  
42

43 Institution: 1. London School of Hygiene and Tropical Medicine, 2. Royal College of Surgeons  
44

45 Reimbursement for attending a symposium?: No  
46

47 A fee for speaking?: No  
48

49 A fee for organising education?: No  
50

51 Funds for research?: No  
52

53 Funds for a member of staff?: No  
54

55 Fees for consulting?: No  
56  
57  
58  
59  
60

1  
2  
3 Have you in the past five years been employed by an organisation that may  
4 in any way gain or lose financially from the publication of this paper?: No  
5

6  
7 Do you hold any stocks or shares in an organisation that may in any way  
8 gain or lose financially from the publication of this paper?: No  
9

10 If you have any competing interests (either as indicated above or any other financial or non-  
11 financial interests) please declare them here:  
12

13  
14 Reviewer: 3  
15

16 Recommendation:  
17

18 Comments:  
19

20 \* Originality -  
21

22 The work highlights opportunities to improve the quality of QI trials  
23

24 \* Importance of work to general readers  
25

26 I believe the work is more valuable to a specialised reader with an interest in methodology.  
27

28 \* Scientific reliability  
29

30 The Research Question is clearly defined and appropriately answered.  
31

32 \* The overall design of study is adequate.  
33

34 It includes recent tools developed specifically for QI studies.  
35

36 \* Participants studied –  
37

38 The full list of studies included in the analysis would have been useful.  
39

40 I reviewed reference 1. to gain a better understanding of the studies reviewed but was unable to  
41 access the online supplementary files.  
42

43 A table listing the QI strategies would be useful.  
44

45 \* Methods –  
46

47 The work complies with relevant reporting standard – SQUIRE and CONSORT.  
48

49 \* Results –  
50

51 The research question is answered.  
52

53  
54 \* Interpretation and conclusions  
55

56 A general reader is likely to have different expectations of QI studies since they reflect real world  
57 conditions where loss to follow up is to be expected.  
58

59 Selective outcome reporting was unclear in 85% of studies. I question the relevance of this  
60

1  
2  
3 measure of bias in a QI study.  
4  
5  
6

7  
8 \* References - up to date and relevant.  
9

10  
11  
12 \*Abstract/summary/key messages - reflect accurately what the paper says.  
13  
14

15 Additional Questions:

16 Please enter your name: Sharon Robyn O'Rourke

17  
18 Job Title: Public Health Physician, Diabetes  
19

20  
21 Institution: Cairns Diabetes Centre  
22

23 Reimbursement for attending a symposium?: No  
24

25 A fee for speaking?: No  
26

27 A fee for organising education?: No  
28

29 Funds for research?: No  
30

31 Funds for a member of staff?: No  
32

33 Fees for consulting?: No  
34

35  
36 Have you in the past five years been employed by an organisation that may  
37 in any way gain or lose financially from the publication of this paper?: No  
38

39 Do you hold any stocks or shares in an organisation that may in any way  
40 gain or lose financially from the publication of this paper?: No  
41

42 If you have any competing interests (either as indicated above or any other financial or non-  
43 financial interests) please declare them here:  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



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

Journal:	<i>BMJ Open</i>
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
<b>Primary Subject Heading</b>:	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

SCHOLARONE™  
Manuscripts

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1           **Quality improvement needed in quality improvement randomised trials:**

2                   **Systematic review of interventions to improve care in diabetes**

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

5   <sup>1</sup>Department of Family and Community Medicine, Women's College Hospital-University  
6   of Toronto, 76 Grenville Street, Toronto, Ontario, M5S1B2, Canada, *Family physician*

7   <sup>2</sup>Li KaShing Knowledge Institute of St Michael's Hospital, 30 Bond Street, Toronto,  
8   Ontario, M5B 1W8, Canada, *Scientist*

9   <sup>3</sup>Clinical Epidemiology Program, Ottawa Hospital Research Institute, 1053 Carling  
10   Avenue, Ottawa, Ontario, K1Y 4E9, Canada, *Scientist*

11   <sup>4</sup>Division of Endocrinology and Metabolism, Department of Medicine, University of  
12   Toronto, 200 Elizabeth St, Toronto, Ontario, M5G2C4, Canada, *Endocrinology fellow*

13   <sup>5</sup>Clinical Epidemiology Program, Ottawa Health Research Institute, Ottawa Hospital -  
14   General Campus, Centre for Practice-Changing Research 501 Smyth Road, Ottawa,  
15   Ontario, K1H 8L6, Canada, *Statistician*

16   <sup>6</sup>Clinical Epidemiology Program, Ottawa Health Research Institute, Ottawa Hospital -  
17   General Campus, Centre for Practice-Changing Research 501 Smyth Road, Ottawa,  
18   Ontario, K1H 8L6, Canada, *Scientist*

19   <sup>7</sup>Clinical Epidemiology Program, Ottawa Health Research Institute, Ottawa Hospital -  
20   General Campus, Centre for Practice-Changing Research 501 Smyth Road, Ottawa,  
21   Ontario, K1H 8L6, Canada, *Senior scientist*



1  
2  
3 22 **Correspondence to:** Noah Ivers, MD, Telephone: 4163236060, ext1 then 2, email:  
4  
5  
6 23 noah.ivers@utoronto.ca, fax: 4163236255.  
7

8  
9 24

10  
11 25 **Ethical approval:** Not required.  
12

13  
14 26 **Data Sharing:** Statistical code and dataset are available from the corresponding author.  
15

16 27 **Word count:** 2797 (main text), 307 (abstract), 25 references, 4 tables, 2 figures.  
17

18 28 **Key words:** systematic review, quality improvement, risk of bias, diabetes mellitus,  
19  
20  
21 29 randomised controlled trials  
22

23  
24 30  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 31 **Article Summary**  
4

5 32 **Article focus**  
6

- 7  
8 33 • Reliable quality improvement research is needed to make decisions about initiating or  
9  
10 34 scaling up quality improvement strategies.  
11  
12 35 • The number of published quality improvement trials has increased rapidly over time.  
13  
14 36 • The quality of trials published in other areas of health seem to be improving over time  
15  
16 37 but the risk of bias in the quality improvement literature is uncertain  
17  
18  
19

20 38 **Key messages**  
21

- 22 39 • Nearly half of quality improvement trials for diabetes are at high risk of bias.  
23  
24 40 • The quality of quality improvement trials does not seem to be improving over time.  
25  
26 41 • Policy-makers, administrators, clinicians, and research funders must carefully  
27  
28 42 scrutinize the methods used in quality improvement trials to ensure evidence-based  
29  
30 43 quality improvement.  
31  
32  
33

34 44 **Strengths and limitations of this study**  
35

- 36  
37 45 • This is the largest systematic review of risk of bias in the quality improvement  
38  
39 46 literature and the only to assess for trends over time.  
40  
41 47 • The risk of bias tool does not capture all sources of methodological bias and poor  
42  
43 48 reporting interferes with the assessment of many domains.  
44  
45 49 • The merits of any given trial report depends to some extent on the needs of the reader,  
46  
47 50 such that some trials with high risk of bias may be of value for certain purposes.  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 **51 Abstract**  
4

5  
6 52 Objective: Despite increasing numbers of published trials of quality improvement (QI)  
7  
8 53 interventions in diabetes, little is known about the risk of bias in this literature.  
9

10  
11 54 Design: Secondary analysis of a systematic review.

12  
13 55 Data sources: Medline, the Cochrane Effective Practice and Organisation of Care (EPOC)  
14  
15 56 database (from inception to July 2010), and references of included studies.  
16

17  
18 57 Eligibility criteria: Randomised trials assessing 11 predefined QI strategies or financial  
19  
20 58 incentives targeting health systems, health-care professionals, or patients to improve  
21  
22 59 management of adult outpatients with diabetes.  
23

24  
25 60 Analysis: Risk of bias (low, unclear, or high) was assessed for the 142 trials in the review  
26  
27 61 across nine domains using the EPOC version of the Cochrane Risk of Bias Tool. We used  
28  
29 62 Cochran-Armitage tests for trends to evaluate improvement over time.  
30

31  
32 63 Results: There was no significant improvement over time in any of the risk of bias  
33  
34 64 domains. Attrition bias (loss to follow up) was the most common source of bias, with 24  
35  
36 65 trials (17%) having high risk of bias due to incomplete outcome data. Overall, 69 trials  
37  
38 66 (49%) had at least one domain with high risk of bias. Inadequate reporting frequently  
39  
40 67 hampered risk of bias assessment: allocation sequence was unclear in 82 trials (58%) and  
41  
42 68 allocation concealment was unclear in 78 trials (55%). There were no significant  
43  
44 69 reductions in the proportions of studies at high risk of bias over time, nor in the adequacy  
45  
46 70 of reporting of risk of bias domains.  
47  
48

49  
50 71 Conclusion: Nearly half of the included QI trials in this review were judged to have high  
51  
52 72 risk of bias. Such trials have serious limitations that put the findings in question and  
53  
54 73 therefore inhibit evidence-based QI. There is a need to limit the potential for bias when  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

74 conducting QI trials and improve the quality of reporting of QI trials so that stakeholders  
75 have adequate evidence for implementation.  
76

For peer review only

1  
2  
3 774  
5 78 **Introduction**

6  
7  
8  
9 79 There is significant interest in quality improvement (QI) in health care, as evidenced by  
10  
11 80 the rapidly increasing number of randomised clinical trials (RCTs) of QI interventions,  
12  
13 81 especially in the diabetes literature.<sup>1</sup> RCTs can provide a foundation for making  
14  
15 82 statements regarding causation, but the validity of trials varies widely; trials with  
16  
17 83 adequate allocation concealment and blinding generally produce smaller effect sizes.<sup>2</sup>  
18  
19 84 Since internal validity in QI trials is a necessary precursor for application to other  
20  
21 85 settings,<sup>3</sup> the ‘risk of bias’ of the findings should be assessed to ascertain the utility of the  
22  
23 86 trial results. When an RCT is deemed to have high risk of bias, the study’s findings  
24  
25 87 become questionable.<sup>4</sup>

26  
27  
28  
29  
30  
31 88 Evaluations to assess trends in methodological quality of RCTs have been conducted in  
32  
33 89 many fields of health care,<sup>5</sup> but no previous reviews have assessed risk of bias in QI  
34  
35 90 RCTs or whether risk of bias in QI RCTs has changed over time. Recently, we conducted  
36  
37 91 a systematic review and meta-regression that included 142 RCTs evaluating QI strategies  
38  
39 92 to improve care for patients with diabetes.<sup>1</sup> In this secondary analysis of those data, we  
40  
41 93 aimed to examine the risk of bias of included studies using the Cochrane Risk of Bias  
42  
43 94 tool developed by the Cochrane Effective Practice and Organisation of Care (EPOC)  
44  
45 95 group<sup>6</sup> and determine whether the proportion with high risk of bias decreased over time.  
46  
47 96 We also evaluated trial and publication characteristics that might be associated with high  
48  
49 97 risk of bias. Finally, we assessed whether the adequacy of reporting of risk of bias  
50  
51 98 domains improved over time.  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 99 **METHODS**  
4  
5

6  
7 100 A detailed description of the methods used for searching, screening, and abstracting the  
8  
9 101 relevant data has been published<sup>1</sup> and is briefly summarized here.  
10

11  
12 102 *Search strategy*  
13

14  
15  
16 103 Studies were identified by searching MEDLINE and the Cochrane EPOC database (up to  
17  
18 104 July 2010), and screening references of included RCTs. The search strategy has been  
19  
20 105 previously published<sup>1</sup> and is available upon request.  
21  
22

23  
24 106 *Study selection*  
25

26  
27 107 RCTs examining one of eleven pre-defined QI strategies, and/or financial incentives,  
28  
29 108 targeting health systems and/or healthcare professionals for the management of adult  
30  
31 109 outpatients with diabetes were included. RCTs had to report at least one of the chosen  
32  
33 110 process of care measures (proportion of patients taking acetylsalicylic acid, statins, anti-  
34  
35 111 hypertensive medication, screened for retinopathy, screened for foot abnormalities,  
36  
37 112 monitored for renal function) or intermediate outcomes (glycosylated haemoglobin  
38  
39 113 levels, low-density lipoprotein cholesterol levels, diastolic and systolic blood pressure,  
40  
41 114 proportion of patients with controlled hypertension, proportion of patients who quit  
42  
43 115 smoking) for inclusion.  
44  
45  
46  
47  
48

49 116 *Data abstraction*  
50

51  
52  
53 117 A draft data abstraction form was developed and modified after a training exercise among  
54  
55 118 reviewers. Two reviewers abstracted relevant data for each RCT independently.  
56  
57  
58  
59  
60

1  
2  
3 119 Discrepancies were resolved by discussion or the involvement of a third reviewer.  
4  
5 120 Authors of the included RCTs were contacted to obtain further information for data items  
6  
7  
8 121 requiring clarification. Journal impact factors from journal citation reports (ISI Web of  
9  
10 122 Science, 2009) were obtained. When a journal's ranking was unavailable, we used the  
11  
12 123 impact ranking of the open access SMImago journal and country rank database, if  
13  
14 124 available.<sup>7</sup> This ranking is calculated using a similar formula and is strongly correlated  
15  
16 125 with the journal citation impact factor.<sup>8</sup>  
17  
18  
19

### 20 21 126 *Assessing Risk of Bias*

22  
23  
24 127 As the included trials tested QI interventions, the Cochrane EPOC Risk of Bias Tool<sup>6</sup> was  
25  
26 128 used to assess the risk of bias in each study. The standard Cochrane Risk of Bias Tool  
27  
28 129 includes an assessment of seven domains: sequence generation, allocation concealment,  
29  
30 130 blinding of participants and personnel, blinding of outcome assessment, incomplete  
31  
32 131 outcome data, selective reporting, and other. The Cochrane Handbook<sup>9</sup> provides  
33  
34 132 instructions for making judgments about the specific domains as high, unclear, or low  
35  
36 133 risk. When formulating summary assessments for each trial, classification of a study as  
37  
38 134 "high risk" indicates that bias could have affected the results, while unclear risk of bias  
39  
40 135 indicates that some doubt exists about the results, and low risk of bias indicates that bias  
41  
42 136 is unlikely to affect the results. It has been shown empirically that studies classified as  
43  
44 137 high risk using this tool are more likely to have larger effect sizes.<sup>10</sup>  
45  
46  
47  
48  
49

50  
51 138 The EPOC tool was adapted to account for the unique features of QI trials. (The  
52  
53 139 guidelines for applying the Cochrane EPOC tool are summarized in Table 1.) For  
54  
55 140 example, in many QI trials it is not possible to blind participants. In addition, QI trials

1  
2  
3 141 may require cluster-randomization to avoid contamination, but in cluster-randomised  
4  
5 142 trials balance at baseline is a particular concern.<sup>11</sup> Therefore, the EPOC tool uses the  
6  
7  
8 143 same approach as the general Cochrane Risk of Bias Tool, but requires an assessment of  
9  
10 144 bias in nine domains: sequence generation, allocation concealment, similarity of baseline  
11  
12 145 measurements, similarity of baseline characteristics, incomplete data, blinding of  
13  
14 146 outcome assessment, contamination, selective outcome reporting, and other. If a given  
15  
16 147 domain is deemed 'unclear' it was inadequately reported to determine whether it meets  
17  
18 148 high risk or low risk criteria. Risk of bias assessment was conducted independently by a  
19  
20 149 clinician-researcher (NMI) and a systematic review methodologist (ACT) and conflicts  
21  
22 150 were resolved by discussion with an expert QI trialist (JMG).  
23  
24  
25  
26  
27

## 28 151 **ANALYSIS**

29  
30  
31 152 For each risk of bias domain, the proportions of RCTs meeting the criteria for high or low  
32  
33 153 or unclear risk of bias were determined. To assess for trends over time in the bias  
34  
35 154 classifications, year of publication was categorized into three groups demarcated by the  
36  
37 155 publication of the 2001 CONSORT statement<sup>12</sup> and the publication of the earlier version  
38  
39 156 of the systematic review of diabetes QI interventions in 2006,<sup>13</sup> as we believed these may  
40  
41 157 have spurred investigators to improve the quality of their trial. Therefore, we categorized  
42  
43 158 year of publication as before 2002; 2002-2006; and 2007-2010. We examined each of the  
44  
45 159 risk of bias domains for change over time descriptively and conducted either exact or  
46  
47 160 asymptotic Cochran-Armitage tests for trend for each item.  
48  
49  
50  
51  
52

53  
54 161 We estimated the proportion of QI RCTs at high risk of bias *overall*, together with 95%  
55  
56 162 asymptotic confidence interval (CI). For this analysis, we created a dichotomous indicator  
57  
58  
59  
60



1  
2  
3 163 for each RCT based on whether or not the study was classified as high risk of bias in at  
4  
5 164 least one domain. To assess for trends in reporting over time, we dichotomized domains  
6  
7  
8 165 as ‘reported’ (low or high risk of bias) and ‘unreported’ (unclear risk of bias). We tested  
9  
10 166 for trend over time in the proportion at high risk of bias overall, hypothesizing that the  
11  
12 167 proportion would decline over time. We used the same year of publication categories and  
13  
14  
15 168 conducted Cochran-Armitage tests for trend of the dichotomous indicator.  
16  
17

18  
19 169 We also conducted a *post-hoc* sensitivity analysis that applied an empirically based rule  
20  
21 170 for assigning high risk of bias overall. Since previous meta-analyses have found that high  
22  
23 171 risk of bias in four specific domains, namely allocation sequence generation, allocation  
24  
25 172 concealment, blinding, and selective outcome reporting are each associated with greater  
26  
27  
28 173 effect size,<sup>22-24</sup> we repeated analyses considering only studies with high risk of bias in  
29  
30 174 these domains as high risk of bias *overall*.  
31  
32

33  
34 175 Finally, we tested for associations between high risk of bias in at least one domain and  
35  
36 176 study characteristics chosen *a priori*: type of diabetes (type 1, type 2, both, unclear), type  
37  
38 177 of allocation (cluster randomised, patient randomised), country (USA or Canada, UK or  
39  
40 178 Western Europe, Other), type of intervention (single, multifaceted), journal impact factor,  
41  
42 179 effective sample size, and year of publication using Chi-squared tests (or Fisher’s exact  
43  
44 180 tests, as appropriate) for categorical and Wilcoxon signed-rank tests for continuous  
45  
46  
47 181 measures. We hypothesized that each of these characteristics may be associated with  
48  
49  
50 182 studies at high risk of bias overall.  
51  
52

53  
54 183 All analyses were conducted in SAS Version 9.2.<sup>14</sup>  
55  
56  
57  
58  
59  
60

1  
2  
3 184 **RESULTS**  
4  
5

6  
7 185 See Figure 1 for a study flow diagram.  
8  
9

10 186 We analyzed 142 studies, with 37 (26%) published before 2002, 46 (32%) between 2002  
11  
12 187 and 2006, and 59 (42%) between 2007 and 2010. These studies evaluated the effects of  
13  
14 188 QI interventions on 123,529 patients with diabetes. Trial and patient characteristics are  
15  
16  
17 189 described in Table 2. The proportions of studies judged to be at low, unclear, or high risk  
18  
19 190 of bias for each domain are illustrated in Figure 2. The domains most commonly at high  
20  
21 191 risk of bias were outcome reporting bias (17%) and similarity across characteristics at  
22  
23 192 baseline (16%). A lack of similarity in outcome measures at baseline (10%), and lack of  
24  
25 193 adequate blinding (8%) were also relatively common domains with high risk of bias.  
26  
27  
28 194 Studies were rarely at high risk of bias due to the allocation sequence generation (4%) or  
29  
30 195 allocation concealment (3%), but these domains were often unclearly reported (57% and  
31  
32 196 55% unclear, respectively). Selective outcome reporting was deemed unclear 84% of the  
33  
34 197 time because published protocols were rarely available and it was often plausible that  
35  
36 198 many more outcomes than those reported were measured. Table 3 indicates a lack of  
37  
38 199 significant trend over time in the proportion of trials at high risk of bias for any given  
39  
40 200 domain. Examination of Table 3 also reveals no trends over time in quality of reporting  
41  
42 201 for any of the risk of bias domains.  
43  
44  
45  
46  
47  
48

49 202 Overall, 48.6% (69/142) of the RCTs had a high risk of bias in at least one domain (95%  
50  
51 203 CI 40.4 to 56.8%). Figure 3 illustrates the rapid increase in number of QI RCTs published  
52  
53 204 over time and the cumulative proportion of trials having at least one domain with high  
54  
55 205 risk of bias up to a given year. In general, the line representing the proportion at high risk  
56  
57  
58  
59  
60

1  
2  
3 206 of bias runs parallel to the number of trials published, consistently accounting for almost  
4  
5 207 half of the studies. Table 4 indicates a lack of significant trend over time in the proportion  
6  
7  
8 208 of trials with at least one domain with high risk of bias: these proportions were 46%,  
9  
10 209 44%, and 54% before 2002, between 2002 and 2006, and after 2006, respectively. Table  
11  
12 210 4 also demonstrates a lack of significant association between any of the study  
13  
14 211 characteristics considered and presence of high risk of bias in at least one domain.  
15  
16  
17  
18 212 The sensitivity analysis, restricting studies defined as high risk of bias overall to those  
19  
20 213 with high risk of bias in one of four domains (allocation sequence generation, allocation  
21  
22 214 concealment, blinding, or selective outcome reporting) also revealed no trends over time  
23  
24  
25 215 – the proportions were 19%, 20%, and 20% before 2002, between 2002 and 2006 and  
26  
27 216 after 2006, respectively (p=0.86).

217

## 218 **DISCUSSION**

### 219 *Main findings*

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

1  
2  
3 227 Our findings indicate that greater efforts are needed to ensure both adequate reporting and  
4  
5 228 methodological conduct of diabetes QI trials.  
6  
7

8  
9 229 We found that poor follow-up, baseline imbalances, and blinding were the most common  
10  
11 230 sources of high risk of bias. Although these domains may be difficult fully control in QI  
12  
13 231 trials, methodological approaches are available to mitigate and/or explore such causes of  
14  
15 232 risk of bias. For example, sensitivity analyses may be used to explore the risk of bias  
16  
17 233 related to loss to follow up, and risk of baseline imbalances in QI trials may be reduced  
18  
19 234 through restricted randomization techniques, especially when trials are cluster-  
20  
21 235 randomized with relatively few clusters. In addition, selective outcome reporting may be  
22  
23 236 limited if more QI trial protocols were registered. Finally, although blinding may be  
24  
25 237 particularly difficult to accomplish in QI trials, this should be clearly reported; if outcome  
26  
27 238 assessment is not blinded, risk of bias could still be limited by using objective outcomes.  
28  
29  
30  
31  
32  
33  
34  
35

### 36 240 *Comparison to literature*

37  
38  
39 241 A systematic review focusing on cluster randomised trials found minimal improvement  
40  
41 242 over time in either reporting or methodological conduct.<sup>17</sup> We found no evidence for a  
42  
43 243 difference in the proportion of cluster-randomised trials at high risk of bias compared to  
44  
45 244 trials in which individuals were allocated. However, imbalance at baseline was a common  
46  
47 245 source of potential bias in diabetes QI trials, possibly owing to inadequate use of  
48  
49 246 restricted randomization in cluster trials.<sup>18</sup> Another systematic review included 35 studies  
50  
51 247 covering a range of health-related fields assessing trends over time in quality criteria for  
52  
53 248 RCTs.<sup>5</sup> Of these, 26 found improvement over time for at least one aspect of  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 249 methodological quality. The domain most commonly noted to have improvement was  
4  
5 250 allocation concealment, but the authors noted that this domain remained either poorly  
6  
7  
8 251 reported or inadequately performed in over half of the examined trials. We found a  
9  
10 252 similarly low proportion of studies clearly reporting adequate allocation concealment, and  
11  
12  
13 253 no evidence of improvement over time.

14  
15  
16 254 Previous authors have noted that QI reports may not contain enough information to  
17  
18 255 inform generalization and allow for replication in different clinical settings.<sup>19</sup> Standards  
19  
20 256 for Quality Improvement Reporting (SQUIRE) guidelines suggest that investigators  
21  
22 257 conducting trials use both SQUIRE and CONSORT to inform their manuscripts.<sup>16</sup>  
23  
24  
25 258 Journal editors should enforce the requirements of both SQUIRE and CONSORT for QI  
26  
27  
28 259 RCTs, possibly by permitting detailed information to be posted as online appendices.  
29  
30 260 Although it might seem onerous to force investigators to address all items in SQUIRE  
31  
32 261 and CONSORT, the risks of poor reporting are substantial. Inadequate description of  
33  
34  
35 262 context could omit essential pre-conditions or important effect modifiers for a successful  
36  
37 263 QI program, while incomplete description of the program itself might lead to failure due  
38  
39  
40 264 to partial implementation.

41  
42  
43 265 *Strengths and limitations*

44  
45  
46 266 To our knowledge, this is the largest analysis of risk of bias ever reported for health care  
47  
48 267 QI RCTs and the only one to assess for trends over time. The findings are strengthened  
49  
50  
51 268 by the rigorous methods used to prepare the data for the systematic review.  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 269 QI evaluations have been criticized based on numerous criteria beyond the risk of bias  
4  
5 270 domains, including short duration of intervention, lack of justification for intervention  
6  
7  
8 271 design, and poor generalizability.<sup>2021</sup> Some important components of methodological  
9  
10 272 quality do not relate to bias (e.g. reporting of a sample size calculation). Thus, it is  
11  
12 273 possible that studies at low risk of bias have important flaws with respect to methodology  
13  
14 274 and/or reporting (and vice-versa), and it is possible that using other scales to assess study  
15  
16 275 quality could have led to different results.<sup>22</sup> While the overall risk of bias assessment  
17  
18 276 using the Cochrane Risk of Bias Tool has been shown to differentiate effect sizes (i.e.,  
19  
20 277 higher risk of bias studies usually have larger effect sizes),<sup>10</sup> studies at high risk of bias  
21  
22 278 may still offer valuable knowledge for QI implementers. The merit of any given report  
23  
24 279 will depend on the needs of the reader, while the current analysis provides an assessment  
25  
26 280 of the progress in the literature as a whole.

27  
28  
29  
30  
31  
32 281 Furthermore, we acknowledge that assigning trials with high risk of bias in a single  
33  
34 282 domain a status of high risk of bias *overall* may be arguable. Nevertheless, our sensitivity  
35  
36 283 analysis led to the same conclusion: there has been no improvement over time in the  
37  
38 284 proportion of trials at high risk of bias in this literature and no particular study  
39  
40 285 characteristics were associated with high risk of bias.

41  
42  
43  
44  
45 286 Another potential limitation stems from our analytical approach regarding change over  
46  
47 287 time; collapsing publication year into three timeframes (pre-2002, 2002-2006, 2007-  
48  
49 288 2010) and testing for trends may have limited our power. These timeframes were chosen  
50  
51 289 *a priori* based on the publication of important documents that we thought might affect the  
52  
53 290 conduct and reporting of these trials. We felt the assumption of linear change over time  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 291 underlying the Cochran-Armitage test for trend was appropriate and in keeping with our  
4  
5 292 hypotheses (e.g. high and unclear risk of bias would decrease gradually over time, while  
6  
7  
8 293 low risk of bias would increase). Risk of type 2 error is tempered by the number of tests  
9  
10 294 performed; the lack of a significant p-value for trend for any level of risk of bias in any  
11  
12 295 domain supports our main conclusion. Finally, this review considered only RCTs from  
13  
14 296 the diabetes literature. It would have been preferable to evaluate a random sample of all  
15  
16 297 QI trials, but adequate QI electronic literature searches have yet to be developed.<sup>25</sup>  
17  
18  
19

20  
21 298 *Implications*  
22

23  
24 299 Published trials testing QI in diabetes are frequently at high risk of bias, producing results  
25  
26 300 that may not be replicable. Clinicians must scrutinize the internal validity of the results as  
27  
28 301 a first step in the process of considering the application of clinical findings for particular  
29  
30 302 patients. Our findings emphasize the need for policy-makers, managers, and/or clinical-  
31  
32 303 administrators seeking to implement QI interventions to apply the same process.<sup>3</sup> It is  
33  
34 304 likely that QI investigators publishing RCTs desire for their work to have a broad impact.  
35  
36 305 To help them accomplish this, research funders and journal editors can play an important  
37  
38 306 role by ensuring that QI trials are reported thoroughly and transparently and are designed  
39  
40 307 in a manner that limits the potential for risk of bias.  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 309 **CONFLICT OF INTEREST:**  
4

5 310 This research received no specific grant from any funding agency in the public,  
6  
7  
8 311 commercial, or not-for-profit sectors. The authors declare that: (i) NM, ACT, MT, IH,  
9  
10 312 LT, DM, & JG received support from Ontario Ministry of Health and Long-term Care,  
11  
12 313 and the Alberta Heritage Foundation for the original systematic review, but the funding  
13  
14 314 agencies had no role in the study design, collection, analysis or interpretation of data,  
15  
16 315 writing of the manuscript or in the decision to submit this manuscript for publication; (ii)  
17  
18 316 NM, ACT, MT, IH, LT, DM, & JG have no relationships with any companies that might  
19  
20 317 have an interest in the submitted work in the previous 3 years; (iii) their spouses,  
21  
22 318 partners, or children have no financial relationships that may be relevant to the submitted  
23  
24 319 work; and (iv) NM, ACT, MT, IH, LT, DM, & JG have no non-financial interests that  
25  
26 320 may be relevant to the submitted work.  
27  
28  
29  
30

31 321 **AUTHOR CONTRIBUTIONS**  
32

33  
34 322 NI and ACT designed and coordinated the study, participated in data collection, data  
35  
36 323 analysis, data interpretation, and drafted the manuscript. MT conducted the analysis and  
37  
38 324 participated in data interpretation and drafting the manuscript. IH, LT, DM, and JMG  
39  
40 325 helped to design the study and write the manuscript. All authors read and approved the  
41  
42 326 final manuscript.  
43  
44

45 327 **DATA SHARING**  
46  
47

48 328 Data are available upon request from the corresponding author.  
49

50 329 **SOURCES OF SUPPORT**  
51

52  
53 330 NMI holds fellowship awards from the Canadian Institutes of Health Research (CIHR)  
54  
55 331 and from the Department of Family and Community Medicine, University of Toronto.  
56  
57  
58  
59  
60



- 1  
2  
3 332 ACT holds a CIHR/Drug Safety and Effectiveness Network new investigator award.  
4  
5  
6 333 JMG and DM both hold Canada Research Chairs.  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

## References

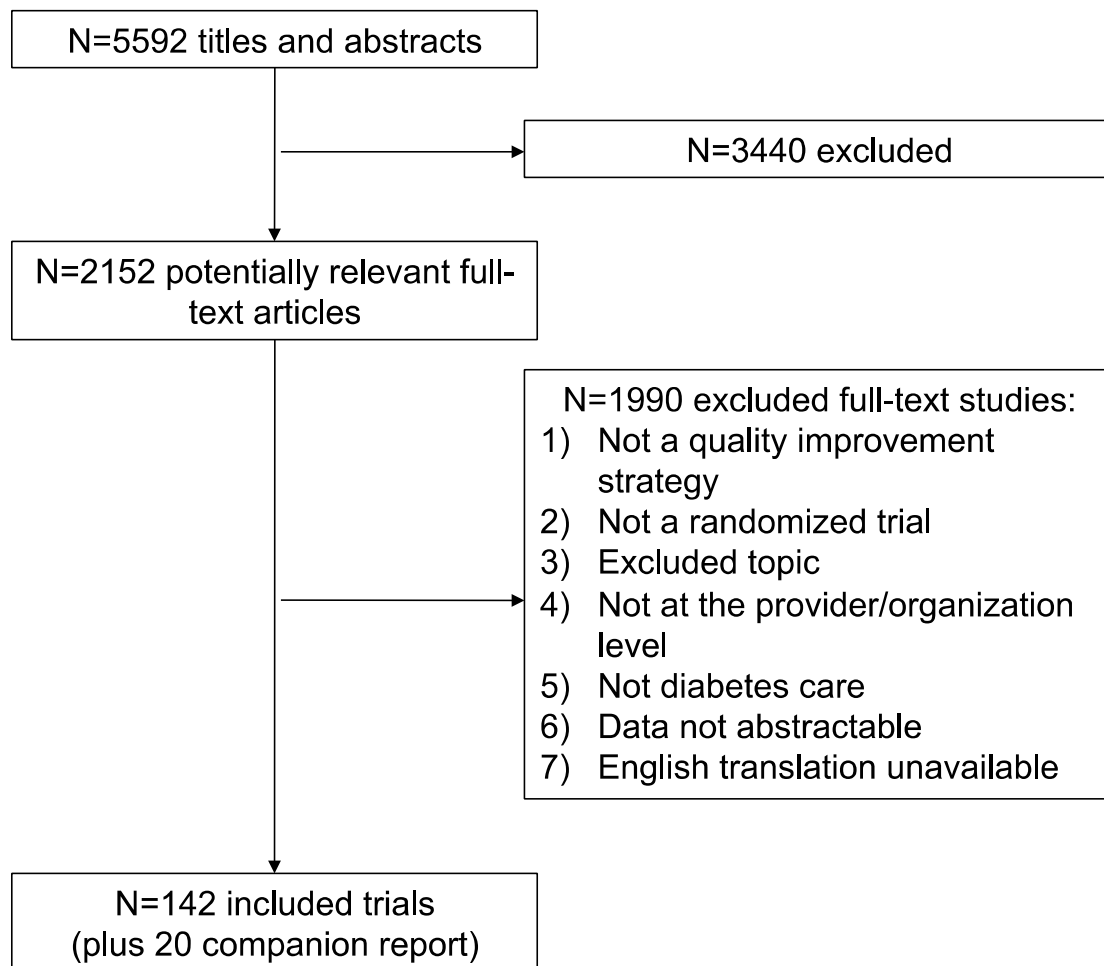
1. Tricco AC, Ivers NM, Grimshaw JM, et al. Effectiveness of quality improvement strategies on the management of diabetes: a systematic review and meta-analysis. *Lancet* 2012;379:2252-2261.
2. Pildal J, Hrobjartsson A, Jorgensen KJ, et al. Impact of allocation concealment on conclusions drawn from meta-analyses 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.

- 1  
2  
3 355 9. Higgins JPT, Altman DG. Assessing risk of bias in included studies. In: Higgins JPT,  
4  
5 356 Green S, ed. *Cochrane handbook for systematic reviews of interventions*: Wiley,  
6  
7 357 2008:187-241.
- 8  
9  
10 358 10. Hartling L, Ospina M, Liang Y, et al. Risk of bias versus quality assessment of  
11  
12 359 randomised controlled trials: cross sectional study. *BMJ* 2009;339:b4012.
- 13  
14  
15 360 11. Ivers NM, Halperin IJ, Barnsley J, et al. Allocation techniques for balance at baseline  
16  
17 361 in cluster randomised trials: a methodological review. *Trials* 2012;13:120.
- 18  
19  
20 362 12. Moher D, Schulz KF, Altman D, et al. The CONSORT statement: revised  
21  
22 363 recommendations for improving the quality of reports of parallel-group randomised trials.  
23  
24 364 *JAMA* 2001;285:1987-1991
- 25  
26  
27 365 13. Shojania KG, Ranji SR, McDonald KM, et al. Effects of quality improvement  
28  
29 366 strategies for type 2 diabetes on glycaemic control: a meta-regression analysis. *JAMA*  
30  
31 367 2006;296:427-440.
- 32  
33  
34 368 14. SAS 9.2 [program]. Cary, NC: SAS Institute Inc., 2008.
- 35  
36  
37 369 15. Simera I, Moher D, Hirst A, et al. Transparent and accurate reporting increases  
38  
39 370 reliability, utility, and impact of your research: reporting guidelines and the EQUATOR  
40  
41 371 Network. *BMC Med* 2010;8:24.
- 42  
43  
44 372 16. Davidoff F, Batalden P, Stevens D, et al. Publication guidelines for quality  
45  
46 373 improvement studies in health care: evolution of the SQUIRE project. *BMJ*  
47  
48 374 2009;338:a3152.
- 49  
50  
51 375 17. Ivers NM, Taljaard M, Dixon S, et al. Impact of CONSORT extension for cluster  
52  
53 376 randomised trials on quality of reporting and study methodology: review of random  
54  
55 377 sample of 300 trials, 2000-8. *BMJ* 2011;343:d5886.
- 56  
57  
58  
59  
60

- 1  
2  
3 378 18. Campbell MK, Elbourne DR, Altman DG, et al. CONSORT statement: extension to  
4  
5 379 cluster randomised trials. *BMJ* 2004;328(7441):702-708.  
6  
7  
8 380 19. Michie S, Fixsen D, Grimshaw JM, et al. Specifying and reporting complex  
9  
10 381 behaviour change interventions: the need for a scientific method. *Implement Sci*  
11  
12 382 2009;4:40.  
13  
14  
15 383 20. Alexander JA, Hearld LR. What can we learn from quality improvement research? A  
16  
17 384 critical review of research methods. *Med Care Res Rev* 2009;66:235-271.  
18  
19  
20 385 21. Eccles M, Grimshaw J, Walker A, et al. Changing the behavior of healthcare  
21  
22 386 professionals: the use of theory in promoting the uptake of research findings. *J Clin*  
23  
24 387 *Epidemiol* 2005;58:107-112.  
25  
26  
27 388 22. Armijo-Olivo S, Stiles CR, Hagen NA, et al. Assessment of study quality for  
28  
29 389 systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and  
30  
31 390 the Effective Public Health Practice Project Quality Assessment Tool: methodological  
32  
33 391 research. *J Eval Clin Pract* 2012;18:12-18  
34  
35  
36 392 23. Odgaard-Jensen J, Vist GE, Timmer A, et al. Randomisation to protect against  
37  
38 393 selection bias in healthcare trials. *Cochrane Database Syst Rev* 2011:MR000012.  
39  
40  
41 394 24. Hempel S, Suttorp MJ, Miles JNV, et al. Empirical Evidence of Associations  
42  
43 395 Between Trial Quality and Effect Size. Rockville (MD): Agency for Healthcare Research  
44  
45 396 and Quality; 2011.  
46  
47  
48 397 25. Hempel S, Rubenstein LV, Shanman RM, et al. Identifying quality improvement  
49  
50 398 intervention publications--a comparison of electronic search strategies. *Implement Sci*  
51  
52 399 2011;6:85.  
53  
54  
55  
56  
57  
58  
59  
60

400 **Figure 1: Study flow diagram**

401



402

22

403

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

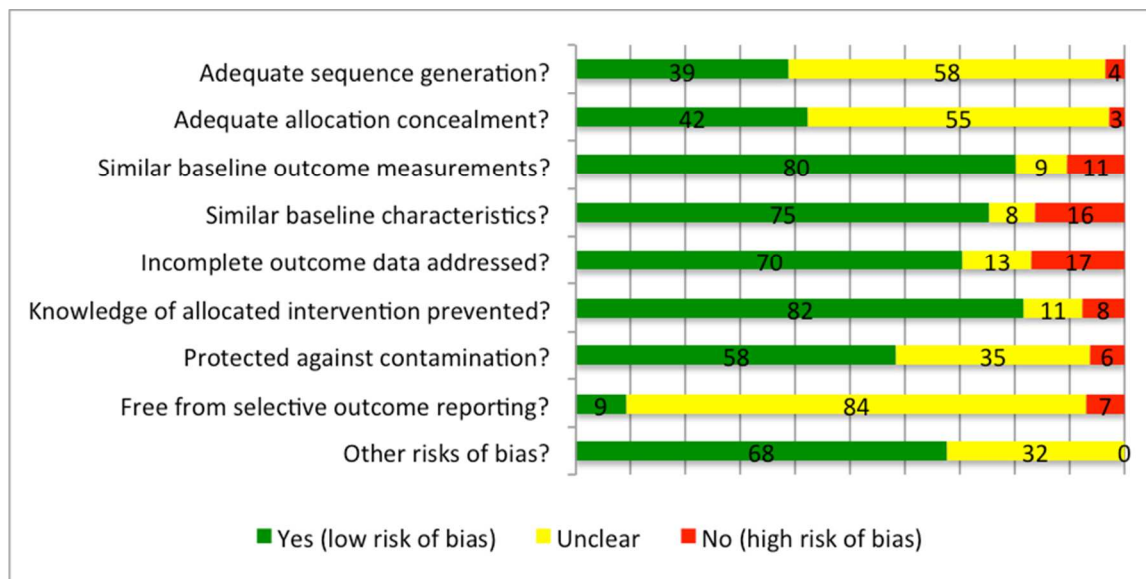
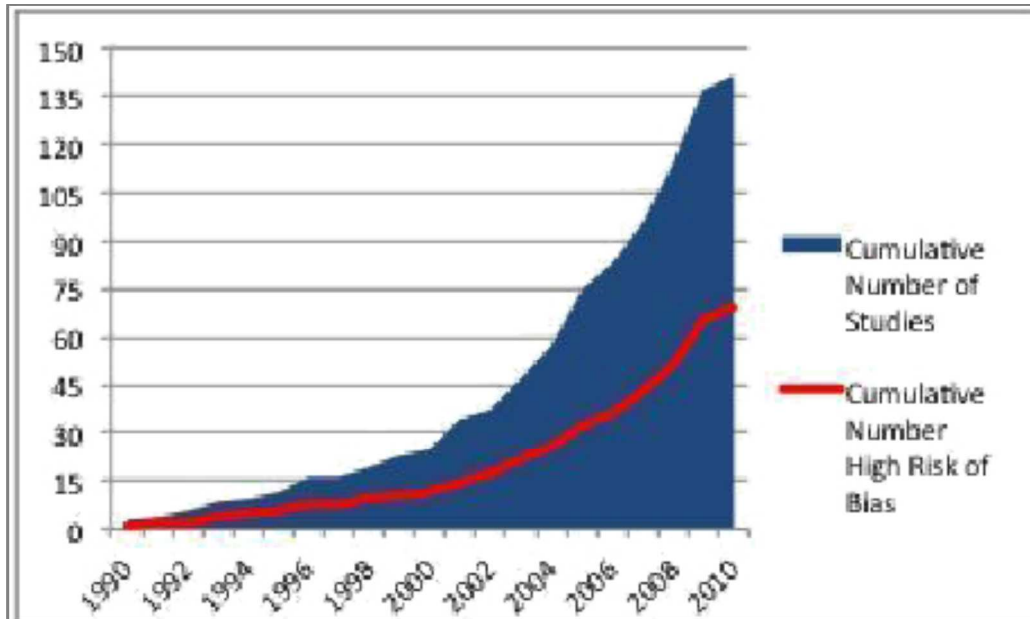


Figure 3: cumulative number of diabetes quality improvement trial publications at high risk of bias in any domain, 1990 to 2010.



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

<b>Risk of Bias Domain</b>	<b>Low Risk of Bias</b>	<b>High Risk of Bias</b>	<b>Unclear Risk of Bias</b>
<b>Was the allocation sequence adequately generated?</b>	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
<b>Was the allocation adequately concealed?</b>	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
<b>Were baseline outcome measurements similar?</b>	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
<b>Were baseline characteristics similar?</b>	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
<b>Were incomplete outcome data adequately addressed?</b>	Missing outcome measures were unlikely to bias the results	Missing outcome data was likely to bias the results.	Not specified in the paper
<b>Was knowledge of the allocated interventions adequately prevented?</b>	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.
<b>Was the study free from selective outcome reporting?</b>	There is no evidence that outcomes were selectively reported	Some important outcomes are omitted from the results	Not specified in the paper
<b>Was the study adequately protected against contamination?</b>	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
<b>Was the study free from other risks of bias?</b>	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 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^
<b>Was the allocation sequence adequately generated?</b>				<b>0.41</b>	<b>0.43</b>
Low	11 (30%)	19 (41%)	25 (42%)		
Unclear	24 (65%)	25 (55%)	33 (56%)		
High	2 (5%)	2 (4%)	1 (2%)		
<b>Was the allocation adequately concealed?</b>				<b>1.00</b>	<b>0.82</b>
Low	15 (40%)	20 (44%)	25 (42%)		
Unclear	21 (57%)	25 (54%)	32 (54%)		
High	1 (3%)	1 (2%)	2 (4%)		
<b>Were baseline outcomes similar?</b>				<b>0.87</b>	<b>0.20</b>
Low	31 (84%)	39 (85%)	44 (75%)		
Unclear	2 (5%)	3 (6%)	8 (13%)		
High	4 (11%)	4 (9%)	7 (12%)		
<b>Were baseline characteristics similar?</b>				<b>0.16</b>	<b>0.57</b>
Low	30 (81%)	34 (74%)	43 (73%)		
Unclear	3 (8%)	6 (13%)	3 (5%)		
High	4 (11%)	6 (13%)	13 (22%)		
<b>Were incomplete outcome data adequately addressed?</b>				<b>0.17</b>	<b>0.70</b>
Low	29 (78%)	33 (72%)	38 (64%)		
Unclear	3 (8%)	8 (17%)	7 (12%)		
High	5 (14%)	5 (11%)	14 (24%)		
<b>Was knowledge of the allocated interventions prevented?</b>				<b>0.44</b>	<b>0.61</b>
Low	32 (87%)	38 (83%)	46 (78%)		
Unclear	3 (8%)	5 (11%)	7 (12%)		
High	2 (5%)	3 (6%)	6 (10%)		
<b>Was the study protected against contamination?</b>				<b>0.54</b>	<b>0.78</b>
Low	25 (68%)	23 (50%)	35 (59%)		
Unclear	10 (27%)	21 (46%)	19 (32%)		
High	2 (5%)	2 (4%)	5 (9%)		
<b>Was the study free from selective outcome reporting?</b>				<b>0.84</b>	<b>1.00</b>
Low	3 (8%)	4 (9%)	6 (10%)		
Unclear	32 (87%)	37 (80%)	50 (85%)		
High	2 (5%)	5 (11%)	3 (5%)		
<b>Was the study free from other risks of bias?</b>				<b>0.58</b>	<b>0.58</b>
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*
<b>Year of publication</b>			<b>0.37</b>
Pre-2002	37	17 (46%)	
2002-2006	46	20 (44%)	
2007-2010	59	32 (54%)	
<b>Type of diabetes</b>			<b>0.11</b>
Type 1	9	3 (33%)	
Type 2	80	36 (45%)	
Both	34	16 (47%)	
Unclear	19	14 (74%)	
<b>Unit of Allocation</b>			<b>0.24</b>
Patient	94	49 (52%)	
Cluster (e.g. provider/clinic)	48	20 (42%)	
<b>Country/Setting</b>			<b>0.62</b>
USA or Canada	79	41 (52%)	
UK or Western Europe	40	17 (43%)	
Other	23	11 (48%)	
<b>Journal Impact Factor</b>			<b>0.87</b>
Greater than 3 (median)	71	34 (47.9%)	
Less than 3 (median)	71	35 (49.3%)	
<b>Effective Sample Size</b>			<b>0.87</b>
Greater than 154 (median)	71	35 (49.3%)	
Less than 154 (median)	71	34 (47.9%)	
<b>Intervention Type</b>			<b>0.17</b>
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

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

1  
2  
3  
4  
5  
6  
7  
8 | **Quality improvement needed in quality improvement ~~randomized~~andomised trials:**

9  
10 | **Systematic review of interventions to improve care in diabetes**

11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
Noah M Ivers<sup>1</sup>, Andrea C Tricco<sup>2</sup>, Monica Taljaard<sup>3</sup>, Ilana Halperin<sup>4</sup>, Lucy Turner<sup>5</sup>,  
David Moher<sup>6</sup>, Jeremy M Grimshaw<sup>7</sup>

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

<sup>2</sup>Li KaShing Knowledge Institute of St Michael's Hospital, 30 Bond Street, Toronto,  
Ontario, M5B 1W8, Canada, *Scientist*

<sup>3</sup>Clinical Epidemiology Program, Ottawa Hospital Research Institute, 1053 Carling  
Avenue, Ottawa, Ontario, K1Y 4E9, Canada, *Scientist*

<sup>4</sup>Division of Endocrinology and Metabolism, Department of Medicine, University of  
Toronto, 200 Elizabeth St, Toronto, Ontario, M5G2C4, Canada, *Endocrinology fellow*

<sup>5</sup>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*

<sup>6</sup>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*

<sup>7</sup>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*

1  
2  
3  
4  
5  
6  
7  
8 22 **Correspondence to:** Noah Ivers, MD, Telephone: 4163236060, ext1 then 2, email:

9  
10 23 noah.ivers@utoronto.ca, fax:4163236255.  
11  
12 24

13  
14  
15 25 **Ethical approval:** Not required.

16  
17 26 **Data Sharing:** Statistical code and dataset are available from the corresponding author.

18  
19 27 **Word count:** 2532 (main text), 307 (abstract), 25 references, 4 tables, 2 figures.

20  
21 28 **Key words:** systematic review, quality improvement, risk of bias, diabetes mellitus,

22  
23 29 ~~randomized~~ ~~andomised~~ controlled trials  
24  
25  
26 30

1  
2  
3  
4  
5  
6  
7  
8 31 **Article Summary**  
9

10 32 **Article focus**

- 11  
12 33 • Reliable quality improvement research is needed to make decisions about initiating or  
13  
14 34 scaling up quality improvement strategies.  
15  
16 35 • The number of published quality improvement trials has increased rapidly over time.  
17  
18 36 • The quality of trials published in other areas of health seem to be improving over time  
19  
20 37 but the risk of bias in the quality improvement literature is uncertain  
21

22 38 **Key messages**

- 23  
24 39 • Nearly half of quality improvement trials for diabetes are at high risk of bias.  
25  
26 40 • The quality of quality improvement trials does not seem to be improving over time.  
27  
28 41 • Policy-makers, administrators, clinicians, and research funders must carefully  
29  
30 42 scrutinize the methods used in quality improvement trials to ensure evidence-based  
31  
32 43 quality improvement.  
33

34 44 **Strengths and limitations of this study**

- 35  
36 45 • This is the largest systematic review of risk of bias in the quality improvement  
37  
38 46 literature and the only to assess for trends over time.  
39  
40 47 • The risk of bias tool does not capture all sources of methodological bias and poor  
41  
42 48 reporting interferes with the assessment of many domains.  
43  
44 49 • The merits of any given trial report depends to some extent on the needs of the reader,  
45  
46 50 such that some trials with high risk of bias may be of value for certain purposes.  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

51 **Abstract**

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

54 Design: Secondary analysis of a systematic review.

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

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

60 Analysis: The risk of bias (low, unclear, or high) was assessed for the 142 trials in the  
61 review across nine domains using the EPOC version of the Cochrane Risk of Bias Tool.

62 We used Cochran-Armitage tests for trends to evaluate improvement over time.

63 Results: There was no significant improvement over time in any of the risk of bias  
64 domains. Attrition bias (loss to follow up) was the most common source of bias, with 24  
65 trials (17%) having high risk of bias due to incomplete outcome data. Inadequate  
66 reporting frequently hampered risk of bias assessment: allocation sequence was unclear in  
67 82 trials (58%) and allocation concealment was unclear in 78 trials (55%). Overall, 69  
68 trials (49%) had at least one domain with high risk of bias. There were no significant  
69 reductions in the proportions of studies that ~~were at unclear or at~~ high risk of bias over  
70 time, nor in the adequacy of reporting of risk of bias domains.

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



1  
2  
3  
4  
5  
6  
7  
8  
9 74 conducting QI trials and improve the quality of reporting of QI trials so that stakeholders  
10 75 have adequate evidence for implementation.  
11  
12 76  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

1  
2  
3  
4  
5  
6  
7  
8 779  
10 **Introduction**

11  
12  
13 79 There is significant interest in quality improvement (QI) in health care, as evidenced by  
14  
15 80 the rapidly increasing number of ~~randomized~~andomised clinical trials (RCTs) of QI  
16  
17 81 interventions, especially in the diabetes literature.<sup>1</sup> RCTs can provide a foundation for  
18  
19 82 making statements regarding causation, but the validity of trials varies widely; trials with  
20  
21 83 adequate allocation concealment and blinding generally produce smaller effect sizes.<sup>2</sup>  
22  
23 84 Since internal validity in QI trials is a necessary precursor for application to other  
24  
25 85 settings,<sup>3</sup> the ‘risk of bias’ of the findings should be assessed to ascertain the utility of the  
26  
27 86 trial results. When an RCT is deemed to have high risk of bias, the study’s findings  
28  
29 87 become questionable.<sup>4</sup>

30  
31 88 Evaluations to assess trends in methodological quality of RCTs have been conducted in  
32  
33 89 many fields of health care,<sup>5</sup> but no previous reviews have assessed risk of bias in QI  
34  
35 90 RCTs or whether risk of bias in QI RCTs has changed over time. Recently, we conducted  
36  
37 91 a systematic review and meta-regression that included 142 RCTs evaluating QI strategies  
38  
39 92 to improve care for patients with diabetes.<sup>1</sup> In this secondary analysis of those data, we  
40  
41 93 aimed to examine the risk of bias of included studies using the Cochrane Risk of Bias  
42  
43 94 tool developed by the Cochrane Effective Practice and Organisation of Care (EPOC)  
44  
45 95 group<sup>6</sup> and determine whether the proportion with high risk of bias decreased over time.  
46  
47 96 We also evaluated trial and publication characteristics that might be associated with high  
48  
49 97 risk of bias. Finally, we assessed whether the adequacy of reporting of risk of bias  
50  
51 98 domains improved over time.

52  
53 6  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8 99 **METHODS**

10  
11 A detailed description of the methods used for searching, screening, and abstracting the  
12  
13 relevant data has been published<sup>1</sup> and is briefly summarized here.

14  
15  
16 102 *Search strategy*

17  
18 103 Studies were identified by searching MEDLINE and the Cochrane EPOC database (up to  
19  
20 104 July 2010), and screening references of included RCTs. The search strategy has been  
21  
22 105 previously published<sup>1</sup> and is available upon request.

23  
24  
25 106 *Study selection*

26  
27  
28 107 RCTs examining one of eleven pre-defined QI strategies, and/or financial incentives,  
29  
30 108 targeting health systems and/or healthcare professionals for the management of adult  
31  
32 109 outpatients with diabetes were included. RCTs had to report at least one of the chosen  
33  
34 110 process of care measures (proportion of patients taking acetylsalicylic acid, statins, anti-  
35  
36 111 hypertensive medication, screened for retinopathy, screened for foot abnormalities,  
37  
38 112 monitored for renal function) or intermediate outcomes (glycosylated haemoglobin  
39  
40 113 levels, low-density lipoprotein cholesterol levels, diastolic and systolic blood pressure,  
41  
42 114 proportion of patients with controlled hypertension, proportion of patients who quit  
43  
44 115 smoking) for inclusion.

45  
46 116 *Data abstraction*

47  
48  
49 117 A draft data abstraction form was developed and modified after a training exercise among  
50  
51 118 reviewers. Two reviewers abstracted relevant data for each RCT independently.

1  
2  
3  
4  
5  
6  
7  
8  
9 119 Discrepancies were resolved by discussion or the involvement of a third reviewer.  
10  
11 120 Authors of the included RCTs were contacted to obtain further information for data items  
12  
13 121 requiring clarification. Journal impact factors from journal citation reports (ISI Web of  
14  
15 122 Science, 2009) were obtained. When a journal's ranking was unavailable, we used the  
16  
17 123 impact ranking of the open access SMIImago journal and country rank database, if  
18  
19 124 available.<sup>7</sup> This ranking is calculated using a similar formula and is strongly correlated  
20  
21 125 with the journal citation impact factor.<sup>8</sup>

### 22 126 *Assessing Risk of Bias*

23  
24  
25 127 As the included trials tested QI interventions, the Cochrane EPOC Risk of Bias Tool<sup>6</sup> was  
26  
27 128 used to assess the risk of bias in each study. The standard Cochrane Risk of Bias Tool  
28  
29 129 includes an assessment of seven domains: sequence generation, allocation concealment,  
30  
31 130 blinding of participants and personnel, blinding of outcome assessment, incomplete  
32  
33 131 outcome data, selective reporting, and other. The Cochrane Handbook<sup>9</sup> provides  
34  
35 132 instructions for making judgments about the specific domains as high, unclear, or low  
36  
37 133 risk. When formulating summary assessments for each trial, classification of a study as  
38  
39 134 "high risk" indicates that bias could have affected the results, while unclear risk of bias  
40  
41 135 indicates that some doubt exists about the results, and low risk of bias indicates that bias  
42  
43 136 is unlikely to affect the results. It has been shown empirically that studies classified as  
44  
45 137 high risk using this tool are more likely to have larger effect sizes.<sup>10</sup>

46  
47 138 The EPOC tool was adapted to account for the unique features of QI trials. (The  
48  
49 139 guidelines for applying the Cochrane EPOC tool are summarized in Table 1.) For  
50  
51 140 example, in many QI trials it is not possible to blind participants. In addition, QI trials

1  
2  
3  
4  
5  
6  
7  
8  
9 141 may require cluster-randomization to avoid contamination, but in cluster-  
10 142 ~~randomized~~~~andomised~~ trials balance at baseline is a particular concern.<sup>11</sup> Therefore, the  
11  
12 143 EPOC tool uses the same approach as the general Cochrane Risk of Bias Tool, but  
13  
14 144 requires an assessment of bias in nine domains: sequence generation, allocation  
15  
16 145 concealment, similarity of baseline measurements, similarity of baseline characteristics,  
17  
18 146 incomplete data, blinding of outcome assessment, contamination, selective outcome  
19  
20 147 reporting, and other. If a given domain is deemed 'unclear' it was inadequately reported  
21  
22 148 to determine whether it meets high risk or low risk criteria. Risk of bias aAssessment was  
23  
24 149 conducted independently by ~~a a clinician-researcher (NMI) and a systematic review~~~~SR~~  
25  
26 150 methodologist (ACT) ~~and a clinician (NMI)~~ and conflicts were resolved by discussion  
27  
28 151 with an expert QI trialist (JMG).

## 30 152 ANALYSIS

31  
32  
33 153 For each risk of bias domain, the proportions of RCTs meeting the criteria for high or low  
34  
35 154 or unclear risk of bias were determined. To assess for trends over time in the bias  
36  
37 155 classifications, year of publication was categorized into three groups demarcated by the  
38  
39 156 publication of the 2001 CONSORT statement<sup>12</sup> and the publication of the earlier version  
40  
41 157 of the systematic review of diabetes QI interventions in 2006,<sup>13</sup> as we believed these may  
42  
43 158 have spurred investigators to improve the quality of their trial. Therefore, we categorized  
44  
45 159 year of publication as before 2002; 2002-2006; and 2007-2010. We examined each of the  
46  
47 160 risk of bias domains for change over time descriptively and conducted either exact or  
48  
49 161 asymptotic Cochran-Armitage tests for trend for each item. ~~Since the number of studies~~

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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

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

172 | ~~For this reason~~ We also conducted a *post-hoc* sensitivity analysis that applied an  
173 | ~~empirically based rule for assigning high risk of bias overall. n, we assessed trends in~~  
174 | ~~individual domains in addition to the summary score and also conducted a~~ Since *post hoc*  
175 | ~~sensitivity analysis that applied an empirically based rule for assigning high risk of bias~~  
176 | ~~overall.~~ Previous meta-analyses have found that high risk of bias in four specific  
177 | ~~domains, namely allocation sequence generation, allocation concealment, blinding, and~~  
178 | ~~selective outcome reporting are each associated with greater effect size.<sup>22-24</sup> we repeated~~  
179 | ~~analyses considering only studies with high risk of bias in these domains as high risk of~~  
180 | ~~bias overall.~~

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

10

Field Code Changed

Formatted: Font: Italic

1  
2  
3  
4  
5  
6  
7  
8 184 | ~~randomized~~andomised), country (USA or Canada, UK or Western Europe, Other), type of  
9  
10 185 | intervention (single, multifaceted), journal impact factor, effective sample size, and year  
11  
12 186 | of publication using Chi-squared tests (or Fisher's exact tests, as appropriate) for  
13  
14 187 | categorical and Wilcoxon signed-rank tests for continuous measures. We hypothesized  
15  
16 188 | that each of these characteristics may be associated with studies at high risk of bias  
17  
18 189 | overall.

19 20  
21 190 | All analyses were conducted in SAS Version 9.2.<sup>14</sup>

## 22 23 191 | **RESULTS**

24  
25  
26 192 | [See Figure 1 for a study flow diagram.](#)

27  
28  
29 193 | We analyzed 142 studies, with 37 (26%) published before 2002, 46 (32%) between 2002  
30  
31 194 | and 2006, and 59 (42%) between 2007 and 2010. These studies evaluated the effects of  
32  
33 195 | QI interventions on 123,529 patients with diabetes. Trial and patient characteristics are  
34  
35 196 | described in Table 2. The proportions of studies judged to be at low, unclear, or high risk  
36  
37 197 | of bias for each domain are illustrated in Figure 2~~1~~. The domains most commonly at high  
38  
39 198 | risk of bias were outcome reporting bias (17%) and similarity across characteristics at  
40  
41 199 | baseline (16%). A lack of similarity in outcome measures at baseline (10%), and lack of  
42  
43 200 | adequate blinding (8%) were also relatively common domains with high risk of bias.  
44  
45 201 | Studies were rarely at high risk of bias due to the allocation sequence generation (4%) or  
46  
47 202 | allocation concealment (3%), but these domains were often unclearly reported (57% and  
48  
49 203 | 55% unclear, respectively). Selective outcome reporting was deemed unclear 84% of the  
50  
51 204 | time because published protocols were rarely available and it was often plausible that

1  
2  
3  
4  
5  
6  
7  
8  
9 205 many more outcomes than those reported were measured. Table 3 indicates a lack of  
10 206 significant trend over time in the proportion of trials at ~~low versus unclear or~~ high risk of  
11  
12 207 bias for any given domain. Examination of Table 3 also reveals no trends over time in  
13  
14 208 quality of reporting for any of the risk of bias domains.

15  
16  
17 209 Overall, 48.6% (69/142) of the RCTs had a high risk of bias in at least one domain (95%  
18  
19 210 CI 40.4 to 56.8%). Figure ~~32~~ illustrates the rapid increase in number of QI RCTs  
20  
21 211 published over time and the cumulative proportion of trials having at least one domain  
22  
23 212 with high risk of bias up to a given year. In general, the line representing the proportion at  
24  
25 213 high risk of bias runs parallel to the number of trials published, consistently accounting  
26  
27 214 for almost half of the studies. Table 4 indicates a lack of significant trend over time in the  
28  
29 215 proportion of trials with at least one domain with high risk of bias: these proportions were  
30  
31 216 46%, 44%, and 54% before 2002, between 2002 and 2006, and after 2006, respectively.

32 217 Table 4 also demonstrates a lack of significant association between any of the study  
33  
34 218 characteristics considered and presence of high risk of bias in at least one domain.  
35  
36  
37 219 The sensitivity analysis, restricting studies defined as high risk of bias overall to those  
38  
39 220 with high risk of bias in one of four domains (allocation sequence generation, allocation  
40  
41 221 concealment, blinding, or selective outcome reporting) also revealed no trends over time  
42  
43 222 – the proportions were 19%, 20%, and 20% before 2002, between 2002 and 2006 and  
44  
45 223 after 2006, respectively (p=0.86).

46  
47  
48 224 ~~Table 3 also demonstrates a lack of significant association between any of the study~~  
49  
50 225 ~~characteristics considered and presence of high risk of bias in at least one domain.~~



1  
2  
3  
4  
5  
6  
7  
8 226 **DISCUSSION**  
9

10  
11 227 *Main findings*  
12

13  
14 228 Using the Cochrane EPOC Risk of Bias Tool,<sup>6</sup> we found that nearly half of RCTs  
15  
16 229 focusing on diabetes had at least one domain at high risk of bias. The trials were most  
17  
18 230 often at high risk of bias due to inadequate follow-up of participants, a lack of similarity  
19  
20 231 at baseline across outcome measures or covariates, or inadequate blinding. We also noted  
21  
22 232 that the majority of RCT reports failed to include an adequate description of the  
23  
24 233 allocation process (i.e., sequence generation and allocation concealment were ‘unclear’).  
25  
26 234 To be interpreted appropriately, RCTs must be completely and transparently reported.<sup>15,16</sup>  
27  
28 235 Our findings indicate that greater efforts are needed to ensure both adequate reporting and  
29  
30 236 methodological conduct of diabetes QI trials.

31  
32 237 We found that poor follow-up, baseline imbalances, and blinding were the most common  
33  
34 238 sources of high risk of bias. Although these domains may be difficult fully control in QI  
35  
36 239 trials, methodological approaches are available to mitigate and/or explore such causes of  
37  
38 240 risk of bias. For example, sensitivity analyses may be used to explore the risk of bias  
39  
40 241 related to loss to follow up, and risk of baseline imbalances in QI trials may be reduced  
41  
42 242 through restricted randomization techniques, especially when trials are cluster-  
43  
44 243 randomized with relatively few clusters. In addition, selective outcome reporting may be  
45  
46 244 limited if more QI trial protocols were registered. Finally, although blinding may be  
47  
48 245 particularly difficult to accomplish in QI trials, this should be clearly reported; if outcome  
49  
50 246 assessment is not blinded, risk of bias could still be limited by using objective outcomes.  
51

1  
2  
3  
4  
5  
6  
7  
8 247 | ~~For example, although blinding may be particularly difficult to accomplish in QI trials,~~  
9 | ~~this should be clearly reported; risk of bias could still be limited by using objective~~  
10 | ~~outcomes.~~  
11  
12  
13

14  
15 250 | *Comparison to literature*

16  
17  
18 251 | A systematic review focusing on cluster ~~randomized~~andomised trials found minimal  
19 | improvement over time in either reporting or methodological conduct.<sup>17</sup> We found no  
20 | evidence for a difference in the proportion of cluster-~~randomized~~andomised trials at high  
21 | risk of bias compared to trials in which individuals were allocated. However, imbalance  
22 | at baseline was a common source of potential bias in diabetes QI trials, possibly owing to  
23 | inadequate use of restricted randomization in cluster trials.<sup>18</sup> Another systematic review  
24 | included 35 studies covering a range of health-related fields assessing trends over time in  
25 | quality criteria for RCTs.<sup>5</sup> Of these, 26 found improvement over time for at least one  
26 | aspect of methodological quality. The domain most commonly noted to have  
27 | improvement was allocation concealment, but the authors noted that this domain  
28 | remained either poorly reported or inadequately performed in over half of the examined  
29 | trials. We found a similarly low proportion of studies clearly reporting adequate  
30 | allocation concealment, and no evidence of improvement over time.  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42

43 264 | Previous authors have noted that QI reports may not contain enough information to  
44 | inform generalization and allow for replication in different clinical settings.<sup>19</sup> Standards  
45 | for Quality Improvement Reporting (SQUIRE) guidelines suggest that investigators  
46 | conducting trials use both SQUIRE and CONSORT to inform their manuscripts.<sup>16</sup>  
47  
48  
49  
50  
51 268 | Journal editors should enforce the requirements of both SQUIRE and CONSORT for QI

1  
2  
3  
4  
5  
6  
7  
8 269 RCTs, possibly by permitting detailed information to be posted as online appendices.  
9

10 270 Although it might seem onerous to force investigators to address all items in SQUIRE  
11 and CONSORT, the risks of poor reporting are substantial. Inadequate description of  
12 context could omit essential pre-conditions or important effect modifiers for a successful  
13 QI program, while incomplete description of the program itself might lead to failure due  
14 to partial implementation.  
15  
16 273  
17  
18 274  
19

20  
21 275 *Strengths and limitations*  
22

23  
24 276 To our knowledge, this is the largest analysis of risk of bias ever reported for health care  
25  
26 277 QI RCTs and the only one to assess for trends over time. The findings are strengthened  
27  
28 278 by the rigorous methods used to prepare the data for the systematic review.  
29

30 279 QI evaluations have been criticized based on numerous criteria beyond the risk of bias  
31  
32 280 domains, including short duration of intervention, lack of justification for intervention  
33  
34 281 design, and poor generalizability.<sup>20,21</sup> Some important components of methodological  
35  
36 282 quality do not relate to bias (e.g. reporting of a sample size calculation). Thus, it is  
37  
38 283 possible that studies at low risk of bias have important flaws with respect to methodology  
39  
40 284 and/or reporting (and vice-versa), and it is possible that using other scales to assess study  
41  
42 285 quality could have led to different results.<sup>22</sup> While the overall risk of bias assessment  
43  
44 286 using the Cochrane Risk of Bias Tool has been shown to differentiate effect sizes (i.e.,  
45  
46 287 higher risk of bias studies usually have larger effect sizes),<sup>10</sup> studies at high risk of bias  
47  
48 288 may still offer valuable knowledge for QI implementers. The merit of any given report  
49  
50 289 will depend on the needs of the reader, while the current analysis provides an assessment  
51  
52 290 of the progress in the literature as a whole.  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9 291 Furthermore, we acknowledge that assigning trials with high risk of bias in a single  
10 292 domain a status of high risk of bias *overall* may be arguable. ~~For this reason, we assessed~~  
11 293 ~~trends in individual domains in addition to the summary score and also conducted a post~~  
12 294 ~~hoc sensitivity analysis that applied an empirically based rule for assigning high risk of~~  
13 295 ~~bias overall. Previous meta-analyses have found that high risk of bias in four specific~~  
14 296 ~~domains, namely allocation sequence generation, allocation concealment, blinding, and~~  
15 297 ~~selective outcome reporting are each associated with greater effect size.~~<sup>23</sup>  
16 298 <sup>24</sup> ~~Nevertheless, our sensitivity analysis considering studies with high risk of bias in any~~  
17 299 ~~of these four (rather than all) domains to be at high risk of bias overall~~ led to the same  
18 300 conclusion: there has been no improvement over time in the proportion of trials at high  
19 301 risk of bias in this literature and no particular study characteristics were associated with  
20 302 high risk of bias.

21  
22  
23  
24  
25  
26 303 Another potential limitation stems from our analytical approach regarding change over  
27 304 time; collapsing publication year into three timeframes (pre-2002, 2002-2006, 2007-  
28 305 2010) and testing for trends may have limited our power. These timeframes were chosen  
29 306 *a priori* based on the publication of important documents that we thought might affect the  
30 307 conduct and reporting of these trials. We felt the assumption of linear change over time  
31 308 underlying the Cochran-Armitage test for trend was appropriate and in keeping with our  
32 309 hypotheses (e.g. high and unclear risk of bias would decrease gradually over time, while  
33 310 low risk of bias would increase). Risk of type 2 error is tempered by the number of tests  
34 311 performed; the lack of a significant p-value for trend for any level of risk of bias in any  
35 312 domain supports our main conclusion. Finally, this review considered only RCTs from

Field Code Changed

1  
2  
3  
4  
5  
6  
7  
8 313 the diabetes literature. It would have been preferable to evaluate a random sample of all  
9 314 QI trials, but adequate QI electronic literature searches have yet to be developed.<sup>25</sup>

10  
11  
12  
13 315 *Implications*

14  
15  
16 316 Published trials testing QI in diabetes are frequently at high risk of bias, producing results  
17 317 that may not be replicable. Clinicians must scrutinize the internal validity of the results as  
18 318 a first step in the process of considering the application of clinical findings for particular  
19 319 patients. Our findings emphasize the need for policy-makers, managers, and/or clinical-  
20 320 administrators seeking to implement QI interventions to apply the same process.<sup>3</sup> It is  
21 321 likely that QI investigators publishing RCTs desire for their work to have a broad impact.  
22 322 To help them accomplish this, research funders and journal editors can play an important  
23 323 role by ensuring that QI trials are reported thoroughly and transparently and are designed  
24 324 in a manner that limits the potential for risk of bias.

25  
26  
27  
28  
29  
30  
31  
32  
33 325  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8 **326 CONFLICT OF INTEREST:**  
9

10 327 This research received no specific grant from any funding agency in the public,  
11 328 commercial, or not-for-profit sectors. The authors declare that: (i) NM, ACT, MT, IH,  
12 329 LT, DM, & JG received support from Ontario Ministry of Health and Long-term Care,  
13 330 and the Alberta Heritage Foundation for the original systematic review, but the funding  
14 331 agencies had no role in the study design, collection, analysis or interpretation of data,  
15 332 writing of the manuscript or in the decision to submit this manuscript for publication; (ii)  
16 333 NM, ACT, MT, IH, LT, DM, & JG have no relationships with any companies that might  
17 334 have an interest in the submitted work in the previous 3 years; (iii) their spouses,  
18 335 partners, or children have no financial relationships that may be relevant to the submitted  
19 336 work; and (iv) NM, ACT, MT, IH, LT, DM, & JG have no non-financial interests that  
20 337 may be relevant to the submitted work.  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30

31 **338 AUTHOR CONTRIBUTIONS**  
32

33 339 NI and ACT designed and coordinated the study, participated in data collection, data  
34 340 analysis, data interpretation, and drafted the manuscript. MT conducted the analysis and  
35 341 participated in data interpretation and drafting the manuscript. IH, LT, DM, and JMG  
36 342 helped to design the study and write the manuscript. All authors read and approved the  
37 343 final manuscript.  
38  
39  
40  
41  
42

43 **344 DATA SHARING**  
44

45 345 Data are available upon request from the corresponding author.  
46

47 **346 SOURCES OF SUPPORT**  
48

49 347 NMI holds fellowship awards from the Canadian Institutes of Health Research (CIHR)  
50 348 and from the Department of Family and Community Medicine, University of Toronto.  
51  
52

1  
2  
3  
4  
5  
6  
7  
8 349 ACT holds a CIHR/Drug Safety and Effectiveness Network new investigator award.  
9

10 350 JMG and DM both hold Canada Research Chairs.  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

351

**References**

- 352 1. Tricco AC, Ivers NM, Grimshaw JM, et al. Effectiveness of quality improvement  
353 strategies on the management of diabetes: a systematic review and meta-analysis. *Lancet*  
354 2012;379:2252-2261.
- 355 2. Pildal J, Hrobjartsson A, Jorgensen KJ, et al. Impact of allocation concealment on  
356 conclusions drawn from meta-analyses of ~~randomized~~~~andomised~~ trials. *Int J Epidemiol*  
357 2007;36:847-857.
- 358 3. Fan E, Laupacis A, Pronovost PJ, et al. How to use an article about quality  
359 improvement. *JAMA* 2010;304:2279-2287.
- 360 4. Higgins JP, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for  
361 assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.
- 362 5. Falagas ME, Grigori T, Ioannidou E. A systematic review of trends in the  
363 methodological quality of ~~randomized~~~~andomised~~ controlled trials in various research  
364 fields. *J Clin Epidemiol* 2009;62:227-231.
- 365 6. Cochrane Effective Practice and Organisation of Care Group. Suggested risk of bias  
366 criteria for EPOC reviews, 2012.  
367 [http://epoc.cochrane.org/sites/epoc.cochrane.org/files/uploads/Suggested%20risk%20of%  
368 20bias%20criteria%20for%20EPOC%20reviews.pdf](http://epoc.cochrane.org/sites/epoc.cochrane.org/files/uploads/Suggested%20risk%20of%20bias%20criteria%20for%20EPOC%20reviews.pdf)
- 369 7. SCImago. SJR — SCImago Journal & Country Rank, 2007.
- 370 8. Falagas ME, Kouranos VD, Arencibia-Jorge R, et al. Comparison of SCImago journal  
371 rank indicator with journal impact factor. *FASEB J* 2008;22:2623-2628.



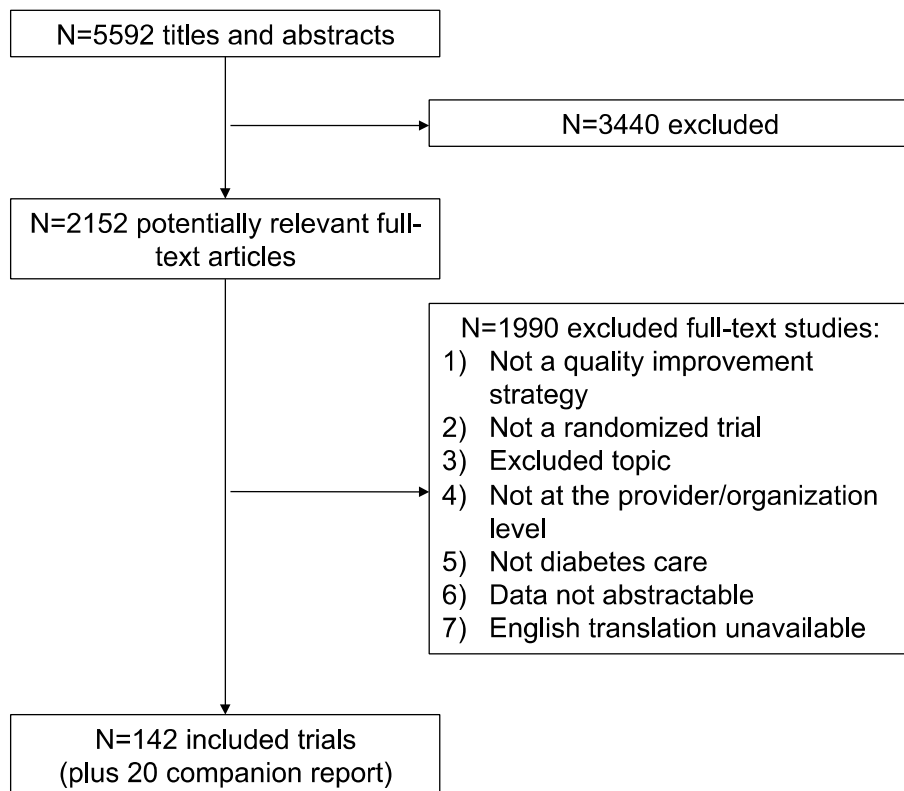
- 1  
2  
3  
4  
5  
6  
7  
8 372 9. Higgins JPT, Altman DG. Assessing risk of bias in included studies. In: Higgins JPT,  
9 373 Green S, ed. *Cochrane handbook for systematic reviews of interventions*: Wiley,  
10 374 2008:187-241.
- 11  
12  
13  
14 375 10. Hartling L, Ospina M, Liang Y, et al. Risk of bias versus quality assessment of  
15 376 randomised controlled trials: cross sectional study. *BMJ* 2009;339:b4012.
- 16  
17  
18 377 11. Ivers NM, Halperin IJ, Barnsley J, et al. Allocation techniques for balance at baseline  
19 378 | in cluster ~~randomized~~andomised trials: a methodological review. *Trials* 2012;13:120.
- 20  
21  
22 379 12. Moher D, Schulz KF, Altman D, et al. The CONSORT statement: revised  
23 380 recommendations for improving the quality of reports of parallel-group  
24 381 | ~~randomized~~andomised trials. *JAMA* 2001;285:1987-1991
- 25  
26  
27 382 13. Shojania KG, Ranji SR, McDonald KM, et al. Effects of quality improvement  
28 383 strategies for type 2 diabetes on glycemic control: a meta-regression analysis. *JAMA*  
29 384 2006;296:427-440.
- 30  
31  
32 385 14. SAS 9.2 [program]. Cary, NC: SAS Institute Inc., 2008.
- 33  
34  
35 386 15. Simera I, Moher D, Hirst A, et al. Transparent and accurate reporting increases  
36 387 reliability, utility, and impact of your research: reporting guidelines and the EQUATOR  
37 388 Network. *BMC Med* 2010;8:24.
- 38  
39  
40 389 16. Davidoff F, Batalden P, Stevens D, et al. Publication guidelines for quality  
41 390 improvement studies in health care: evolution of the SQUIRE project. *BMJ*  
42 391 2009;338:a3152.
- 43  
44  
45 392 17. Ivers NM, Taljaard M, Dixon S, et al. Impact of CONSORT extension for cluster  
46 393 randomised trials on quality of reporting and study methodology: review of random  
47 394 sample of 300 trials, 2000-8. *BMJ* 2011;343:d5886.

- 1  
2  
3  
4  
5  
6  
7  
8 395 18. Campbell MK, Elbourne DR, Altman DG, et al. CONSORT statement: extension to  
9  
10 396 cluster randomised trials. *BMJ* 2004;328(7441):702-708.  
11  
12 397 19. Michie S, Fixsen D, Grimshaw JM, et al. Specifying and reporting complex  
13  
14 398 behaviour change interventions: the need for a scientific method. *Implement Sci*  
15  
16 399 2009;4:40.  
17  
18 400 20. Alexander JA, Heard LR. What can we learn from quality improvement research? A  
19  
20 401 critical review of research methods. *Med Care Res Rev* 2009;66:235-271.  
21  
22 402 21. Eccles M, Grimshaw J, Walker A, et al. Changing the behavior of healthcare  
23  
24 403 professionals: the use of theory in promoting the uptake of research findings. *J Clin*  
25  
26 404 *Epidemiol* 2005;58:107-112.  
27  
28 405 22. Armijo-Olivo S, Stiles CR, Hagen NA, et al. Assessment of study quality for  
29  
30 406 systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and  
31  
32 407 the Effective Public Health Practice Project Quality Assessment Tool: methodological  
33  
34 408 research. *J Eval Clin Pract* 2012;18:12-18  
35  
36 409 23. Odgaard-Jensen J, Vist GE, Timmer A, et al. Randomisation to protect against  
37  
38 410 selection bias in healthcare trials. *Cochrane Database Syst Rev* 2011:MR000012.  
39  
40 411 24. Hempel S, Suttrop MJ, Miles JNV, et al. Empirical Evidence of Associations  
41  
42 412 Between Trial Quality and Effect Size. Rockville (MD): Agency for Healthcare Research  
43  
44 413 and Quality; 2011.  
45  
46 414 25. Hempel S, Rubenstein LV, Shanman RM, et al. Identifying quality improvement  
47  
48 415 intervention publications--a comparison of electronic search strategies. *Implement Sci*  
49  
50 416 2011;6:85.  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

417 **Figure 1: Study flow diagram**

418



Formatted: Font: Bold

419

23

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

420

**Figure 21:** percentage of studies judged to be at low, unclear, or high risk of bias in each risk of bias domain

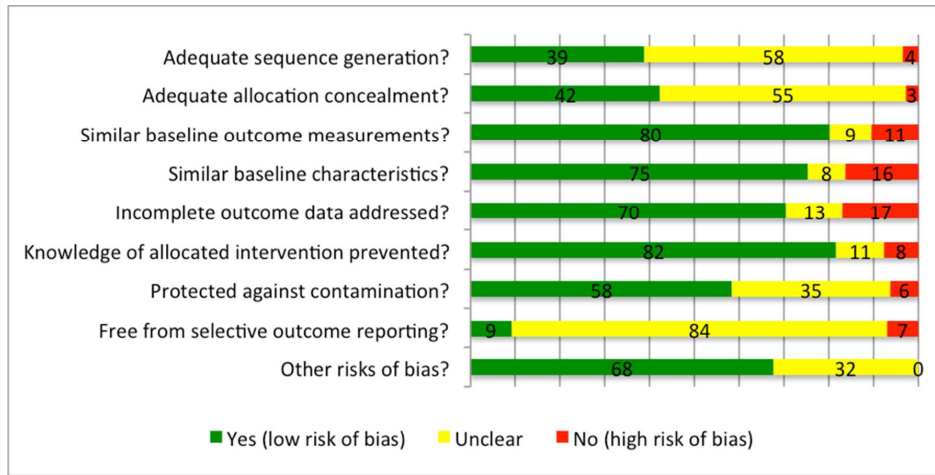


Figure 32: cumulative number of diabetes quality improvement trial publications at high risk of bias in any domain, 1990 to 2010.

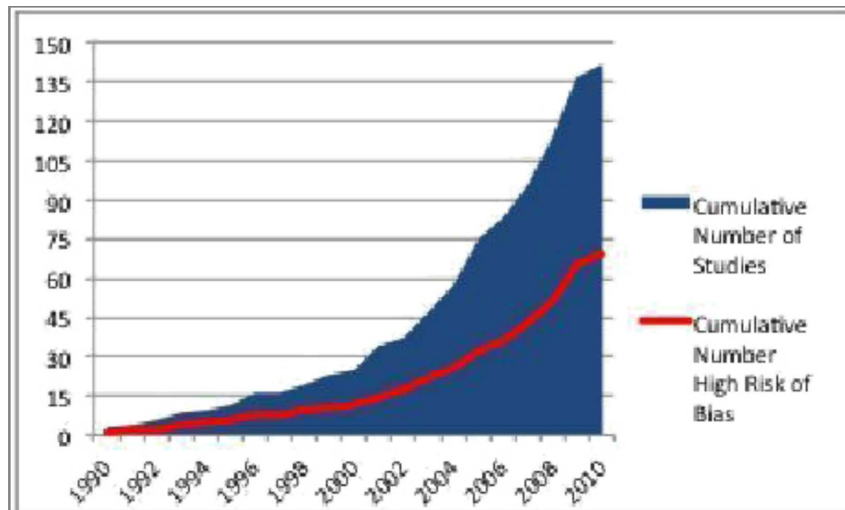


Table 1: Cochrane Effective Practice and organization of care (epoc) risk of bias assessment tool\*

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

Formatted: Line spacing: single



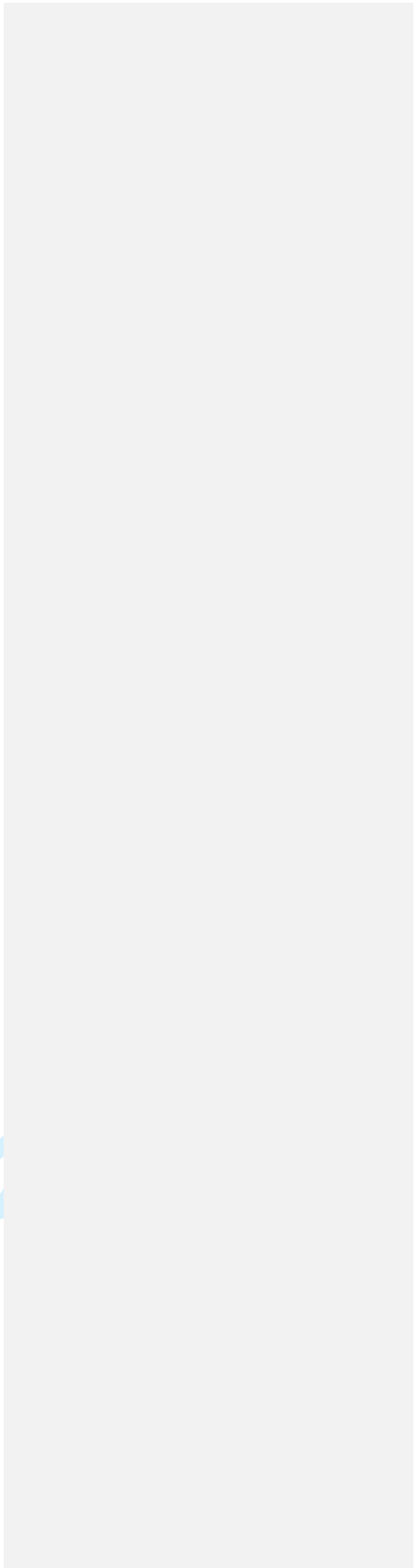
Table 4: association between study characteristics and risk of bias

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

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only





# PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
<b>TITLE</b>			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
<b>ABSTRACT</b>			
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
<b>INTRODUCTION</b>			
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
<b>METHODS</b>			
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

Page 1 of 2

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
<b>FUNDING</b>			
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](http://www.prisma-statement.org).

Page 2 of 2