

## REVIEWS

# Growing Literature, Stagnant Science? Systematic Review, Meta-Regression and Cumulative Analysis of Audit and Feedback Interventions in Health Care

Noah M. Ivers, MD, PhD<sup>1</sup>, Jeremy M. Grimshaw, PhD<sup>2</sup>, Gro Jamtvedt, PT<sup>3</sup>, Signe Flottorp, MD<sup>3</sup>, Mary Ann O'Brien, PhD<sup>1</sup>, Simon D. French, PhD<sup>4</sup>, Jane Young, MD<sup>5</sup>, and Jan Odgaard-Jensen, PhD<sup>3</sup>

<sup>1</sup>Family Practice Health Centre and Institute for Health Systems Solutions and Virtual Care, Women's College Hospital, Toronto, Ontario, Canada; <sup>2</sup>Clinical Epidemiology Program, Ottawa Hospital Research Institute, Department of Medicine, University of Ottawa, Ottawa, Ontario, Canada; <sup>3</sup>Norwegian Knowledge Centre for the Health Services, Oslo, Norway; <sup>4</sup>School of Rehabilitation Therapy, Faculty of Health Sciences, Queen's University, Kingston, Ontario, Canada; <sup>5</sup>Cancer Epidemiology and Services Research, Sydney School of Public Health, University of Sydney, Sydney, New South Wales, Australia.

**BACKGROUND:** This paper extends the findings of the Cochrane systematic review of audit and feedback on professional practice to explore the estimate of effect over time and examine whether new trials have added to knowledge regarding how optimize the effectiveness of audit and feedback.

**METHODS:** We searched the Cochrane Central Register of Controlled Trials, MEDLINE, and EMBASE for randomized trials of audit and feedback compared to usual care, with objectively measured outcomes assessing compliance with intended professional practice. Two reviewers independently screened articles and abstracted variables related to the intervention, the context, and trial methodology. The median absolute risk difference in compliance with intended professional practice was determined for each study, and adjusted for baseline performance. The effect size across studies was recalculated as studies were added to the cumulative analysis. Meta-regressions were conducted for studies published up to 2002, 2006, and 2010 in which characteristics of the intervention, the recipients, and trial risk of bias were tested as predictors of effect size.

**RESULTS:** Of the 140 randomized clinical trials (RCTs) included in the Cochrane review, 98 comparisons from 62 studies met the criteria for inclusion. The cumulative analysis indicated that the effect size became stable in 2003 after 51 comparisons from 30 trials. Cumulative meta-regressions suggested new trials are contributing little further information regarding the impact of common effect modifiers. Feedback appears most effective when: delivered by a supervisor or respected colleague; presented frequently; featuring both specific goals and action-plans; aiming to decrease the targeted behavior; baseline performance is lower; and recipients are non-physicians.

**DISCUSSION:** There is substantial evidence that audit and feedback can effectively improve quality of care, but little evidence of progress in the field. There are opportunity costs for patients, providers, and health care systems when investigators test quality improvement interventions that do not build upon, or contribute toward, extant knowledge.

**KEY WORDS:** audit and feedback; scientific progress; quality improvement; systematic review; cumulative analysis.

J Gen Intern Med 29(11):1534-41

DOI: 10.1007/s11606-014-2913-y

© The Author(s) 2014. This article is published with open access at Springerlink.com

## BACKGROUND

Audit and feedback is widely used as a strategy to improve professional practice, either on its own or as a key component of multifaceted quality improvement (QI) interventions. Providing data regarding clinical performance may overcome health professionals' limited abilities to accurately self-assess their performance.<sup>1</sup> It is posited that when well-designed feedback demonstrates suboptimal performance for important and actionable targets, recipients are more likely to respond with efforts to improve quality of care.<sup>2</sup>

The most recent Cochrane systematic review and meta-analysis of audit and feedback included 140 randomized clinical trials (RCTs),<sup>3</sup> making audit and feedback one of the most studied healthcare quality improvement (QI) interventions. Three Cochrane reviews over the course of 10 years came to the same conclusion: Audit and feedback generally leads to small but potentially important improvements in professional practice (Table 1).<sup>3-5</sup> Yet despite the increasing number of audit and feedback trials (Fig. 1), uncertainty remains regarding when audit and feedback is likely to be most helpful and how to best optimize the intervention.<sup>6</sup>

An abstract of the preliminary findings were presented at the Cochrane Colloquium, October 2011.

Received November 26, 2013

Revised April 7, 2014

Accepted May 16, 2014

Published online June 26, 2014

**Table 1. Findings from Cochrane Systematic Reviews and Meta-Analyses of Audit and Feedback Over Time**

Year of review	Effect size	Conclusion
2003 (search up to January 2001)	Forty-seven studies with dichotomous outcomes: 7 % (IQR: 2–11) median absolute increase in compliance with intended professional behaviors or processes	“Audit and feedback can be effective in improving professional practice. When it is effective, the effects are generally small to moderate. The absolute effects of audit and feedback are more likely to be larger when baseline adherence to recommended practice is low.” <sup>4</sup>
2006 (search up to January 2004)	Forty-nine studies with dichotomous outcomes: 5 % (IQR: 3–11) median absolute increase in compliance with intended professional behaviors or processes	“Audit and feedback can be effective in improving professional practice. The effects are generally small to moderate. The absolute effects are likely to be larger when baseline adherence to recommended practice is low and intensity of audit and feedback is high.” <sup>5</sup>
2012 (search up to December 2010)	Sixty-two studies with dichotomous outcomes: 4 % (IQR: 1–16) weighted median absolute increase in compliance with intended professional behaviors or processes	“Audit and feedback generally leads to small but potentially important improvements in professional practice. The effectiveness of audit and feedback seems to depend on baseline performance and how the feedback is provided. Future studies of audit and feedback should directly compare different ways of providing feedback.” <sup>3</sup>

*IQR interquartile range*

In some instances, audit and feedback is highly effective; learning from such examples is necessary to optimize the effectiveness of the intervention across different contexts. The Cochrane review and associated re-analyses have found that the effectiveness of audit and feedback depends to some extent on how the intervention is designed and delivered, suggesting an opportunity to maximize the impact of this QI strategy on quality of care.<sup>3,7,8</sup> However, there is evidence that many audit and feedback interventions are developed and tested without an explicit attempt to consider relevant theories or to build upon extant knowledge.<sup>9</sup> Ideally, results of early studies would inform the design of future interventions, and through this process, cumulative knowledge would lead to more effective QI. Given the continuing human and financial capital invested in audit and feedback interventions in health care, it is important to examine whether newer trials of audit and feedback have contributed new knowledge to the field.

The purpose of this paper is to extend the results of the Cochrane review of audit and feedback to explore the evolution of evidence supporting this QI intervention over

time. In particular, we examined whether effect estimates, and the precision around those estimates, changed over time. To do this, we undertook a cumulative analysis of trials by year of publication and conducted a series of meta-regressions to understand how the literature has developed with respect to determining factors that could explain why audit and feedback is more or less effective.

**METHODS**

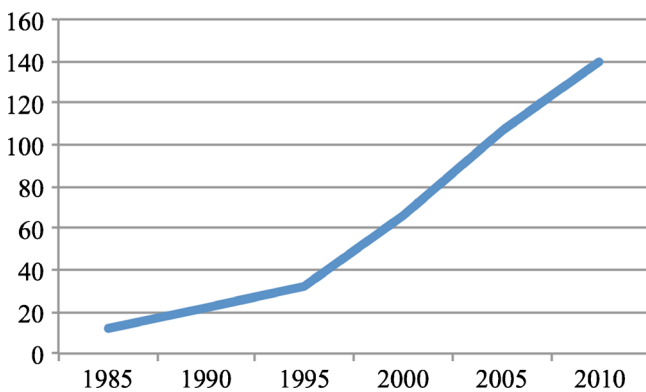
This is a secondary analysis of data from the previously published Cochrane systematic review of audit and feedback. Complete methodological details are available<sup>3</sup> and are summarized below. Ethics approval was not required for this study.

**Eligibility Criteria**

Audit and feedback was defined as a “summary of clinical performance of health care over a specified period of time.” This secondary analysis only included RCTs that directly compared audit and feedback (either alone or as the core, essential feature of a multifaceted intervention) to usual care. Furthermore, only RCTs that evaluated effects on provider practice as a primary outcome were included. For ease of interpretation of the meta-regression and cumulative meta-analysis, we further limited studies to those that reported dichotomous outcomes (i.e., compliance with intended professional practice).

**Information Sources, Search, and Study Selection**

A search strategy sensitive for RCTs involving audit and feedback was applied in December 2010 to the Cochrane Central Register of Controlled Trials, MEDLINE, EMBASE, and CINAHL. As previously described,<sup>3</sup> we



**Figure 1. Cumulative number of randomized trials featuring audit and feedback as a core component of a quality improvement intervention.**

developed a MEDLINE search strategy that identified 89 % of all MEDLINE indexed studies from the previous version of the review and then translated this strategy into the other databases using the appropriate controlled vocabulary as applicable. Search terms included: audit, benchmarking, feedback, utilization review, health care quality, etcetera, plus typical search terms to focus on RCTs. Two reviewers independently screened the titles, abstracts, and full texts to apply inclusion criteria.

## Data Collection Process

Two reviewers independently abstracted data from included studies. Studies included in the previous version of the Cochrane review of audit and feedback were reassessed due to changes in the data abstraction form and methods. Discrepancies were resolved through discussion. For studies lacking extractable data or without baseline information, we contacted investigators via email. Risk of bias for the primary outcome(s) in each study was assessed according to the Cochrane Effective Practice and Organization of Care group criteria<sup>10</sup> (sequence generation, allocation concealment, blinding, incomplete outcome data, selective reporting, baseline similarity, lack of contamination, and other). We assigned an overall assessment of the risk of bias for each study as high, moderate, or low, following the recommendations in the Cochrane Handbook.<sup>11</sup> Studies with a high risk of bias in at least one domain that decreased the certainty of the effect size of the primary outcome were considered to have a high risk of bias. Conversely, when a study had low risk of bias for each domain, it was deemed low risk of bias overall. Other studies were considered to have unclear risk of bias.

## Measure of Treatment Effect

We only extracted results for the primary outcome. When the primary outcome was not specified, we used the variable described in the sample size calculation as the primary outcome. When the primary outcome was still unclear or when the manuscript described several primary process outcomes, we calculated the median value. We calculated the treatment effect as an adjusted risk difference (RD) by subtracting baseline differences from post-intervention differences. Thus, an adjusted RD of +10 % indicates that after accounting for baseline differences, health professionals receiving the intervention adhered to the desired practice 10 % more often than those not receiving the intervention.

## Analysis

Across multiple studies, we weighted the median effect by the number of health care providers. The ‘median of medians’

technique has been used in many similar reviews evaluating the effect of QI interventions on health professional performance,<sup>12</sup> due to frequency of unit of analysis errors in the literature and the great variety of clinical contexts covered in the studies. For the cumulative analysis, the median adjusted RD and interquartile range (IQR) was recalculated at each time point as studies were added. The meta-regression examined how the adjusted RD was related to explanatory variables, weighted according to study size (number of health care professionals). Unlike the meta-regression from the Cochrane review of audit and feedback,<sup>3</sup> high risk of bias studies were included. The meta-regression also tested the following potential sources of heterogeneity to explain variation in the results of the included studies: format (verbal, written, both, unclear); source (supervisor or senior colleague, professional standards review organization or representative of employer/purchaser, investigators, unclear); frequency (weekly, monthly, less than monthly, one-time); instruction for improvement (explicit measurable target or specific goal but no action plan, action plan with suggestions or advice given to help participants improve but no goal/target, both, neither); direction of change required (increase current behavior, decrease current behavior, mix or unclear); recipient (physician, other health professional); and study risk of bias (high, unclear, low). Meta-regression was conducted for all published trials as of 2010, 2006 and 2002. Finally, we added year of publication as a continuous variable to the meta-regression of all studies as an additional approach to assess whether this variable accounted for a significant portion of the heterogeneity. We conducted a multivariable linear regression using main effects only. Baseline compliance and year of publication were treated as continuous explanatory variables and the others as categorical. The analyses were conducted using the GLIMMIX procedure in SAS Version 9.2 (SAS Institute Inc. Cary, NC USA), accounting for the dependency between comparisons from the same trial.

## RESULTS

Of the 140 RCTs included in the Cochrane review, 98 comparisons from 62 studies met the criteria for this study (Fig. 2). These studies included over 2,300 groups of healthcare providers (e.g., clinics or hospitals) from 38 trials allocating clusters of professionals, and more than 2,000 professionals from 24 trials allocating individual healthcare providers. Characteristics of these studies are described in Table 2. Most studies took place in the USA or Canada (55 %), and outpatient care was the most common setting (69 %). The feedback was delivered by either the investigators or by an unclear source in 85 % of studies. In 47 % of the studies, the feedback was only delivered once, and in 61 % the feedback did not include an explicit goal or action plan.

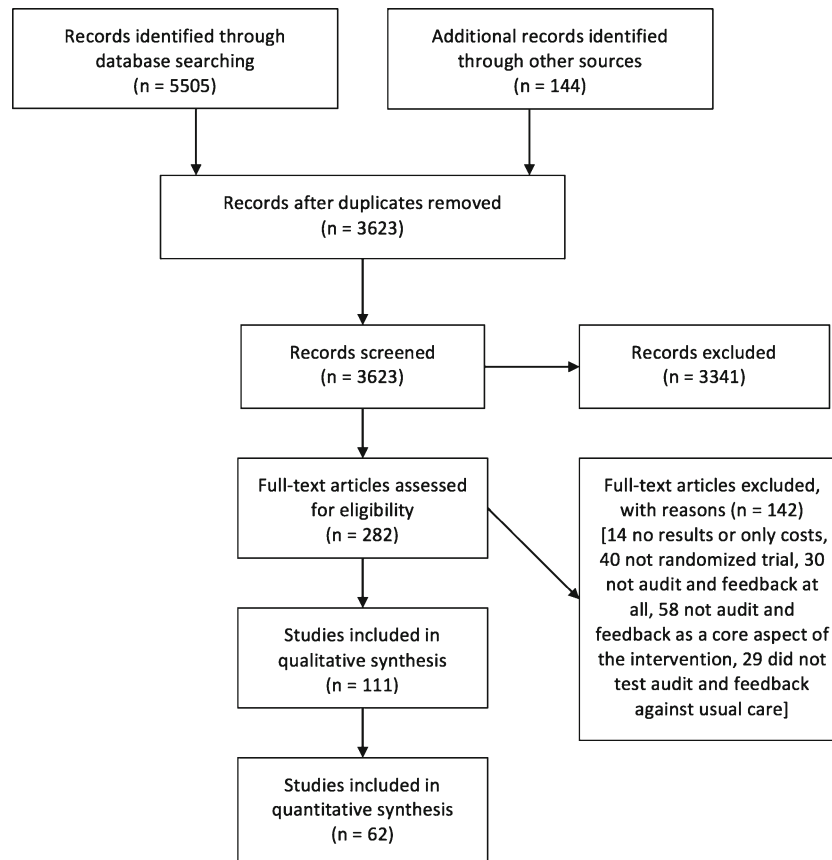


Figure 2. Study flow diagram.

The cumulative analysis revealed little change in the median effect or the interquartile range over the course of 25 years (Fig. 3). The median improvement in adherence to intended practice in 2002 after 51 comparisons had been published was 5.7 (IQR=1.65–10.85), the effect in 2006 after 86 comparisons was 3.5 (IQR=0.65–9.00), and the effect after including all 98 comparisons was 4.4 (IQR=1.04–10.90).

The meta-regression revealed that heterogeneity in effect sizes could be explained in part by feedback characteristics, but year of publication did not explain a significant portion of the variability in effect size (Table 3). Feedback seemed most effective when it was: delivered from a supervisor or respected colleague; presented more than once; featuring both specific goals and action-plans; aiming to decrease the targeted behavior; with lower baseline performance; and when recipients were non-physicians. Studies published after 2006 did not change the meta-regression results statistically; differences in the estimated effect for most feedback characteristics have been apparent qualitatively since 2002. For example, although the p value was not significant for source of feedback in 2002, the estimated adjusted risk difference for feedback delivered by a supervisor or respected colleague (24.5) was higher than feedback delivered by study investigators (17.9) or feed-

back from a representative of a regulatory agency or employer (0.9).

## DISCUSSION

Audit and feedback works; the median effect is small though still potentially important at the population level, and 27/98 comparisons (28 %) resulted in an improvement of at least 10 % in quality of care.<sup>3</sup> Small differences in the results seen in these re-analyses compared to the results of the Cochrane review are due to the lack of weighting in the cumulative analysis and the inclusion of high risk of bias studies in the meta-regression. Nevertheless, the expected effect of an intervention comparing audit and feedback to usual care has changed very little over the last two decades. Furthermore, new trials have provided little new knowledge regarding key effect modifiers. Given the lack of equipoise, it may no longer be ethically appropriate to continue to direct human and financial resources toward trials comparing audit and feedback against usual care, especially for common conditions in common settings. At this point, the appropriate question is not, ‘can audit and feedback improve professional practice?’ but ‘how

Table 2. Characteristics of Studies

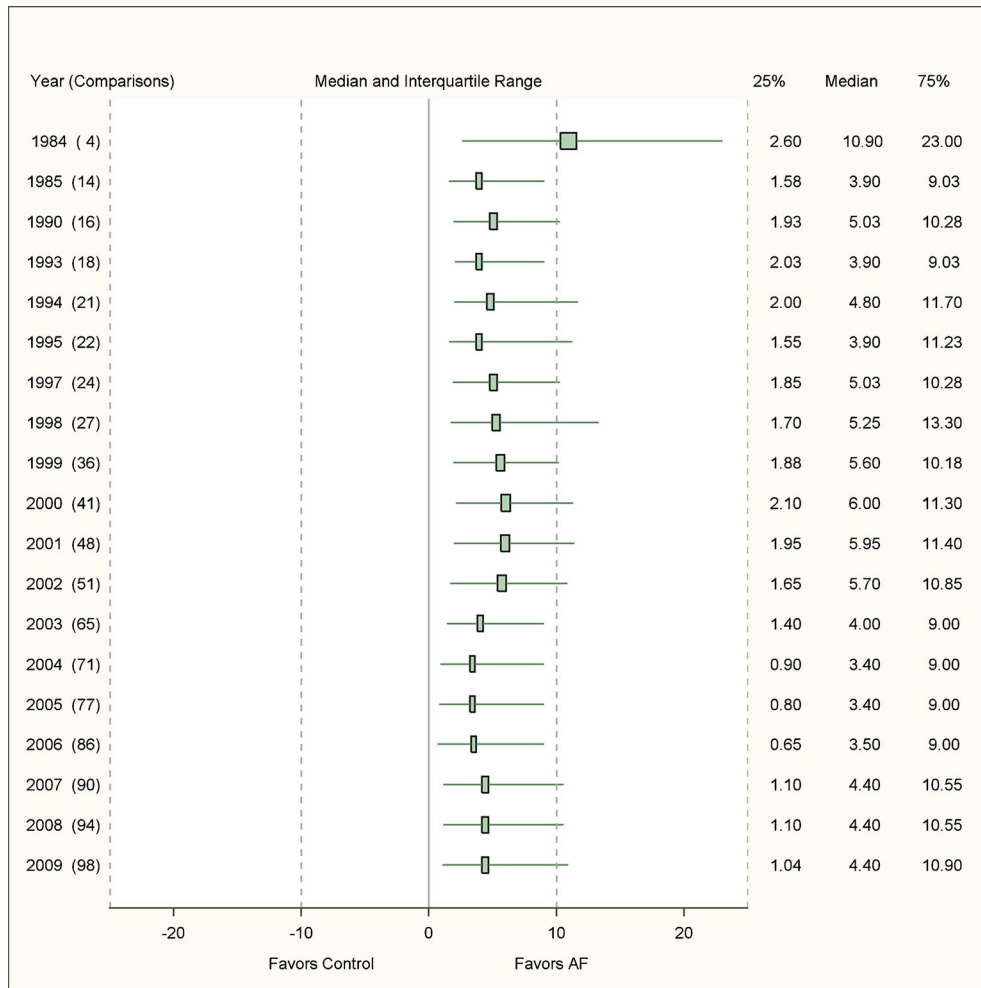
Study characteristic	No.	%	Intervention characteristic	No.	%
Publication year			Format		
2007–2009	10	16.1	Verbal	6	9.7
2003–2006	22	35.5	Written	34	54.8
1996–2002	20	32.4	Both	18	29.0
1986–1995	5	8.1	Unclear	4	6.5
Before 1986	5	8.1	Source		
Country			Supervisor/colleague	6	9.7
USA	27	43.6	Employer	3	4.8
UK or Ireland	7	11.3	Investigators/unclear	53	85.5
Canada	7	11.3	Frequency		
Australia or New Zealand	6	9.7	Monthly or more	11	17.7
Other	15	24.2	Repeated less than monthly	19	30.7
Unit of allocation			Once only	29	46.8
Provider	24	38.7	Instructions for improvement		
Many providers/groups	38	61.3	Goal-setting	5	8.1
Unit of analysis			Action planning	16	25.8
Patient	37	59.7	Both	3	4.8
Provider	12	19.4	Neither	38	61.3
Many providers/groups	12	19.4	Nature of change required		
Unclear	1	1.6	Increase current behavior	29	46.8
Risk of bias			Decrease current behavior	6	9.7
Low	21	33.9	Mix or unclear	27	43.6
Unclear	29	46.8	Clinical topic		
High	12	19.4	Diabetes	11	17.7
Number of arms in trial			Laboratory testing/radiology	3	4.8
Two	41	66.1	Prescribing	18	29.0
Three	10	16.1	Other	30	48.4
Four	11	17.7	Targeted health professional		
Clinical setting			Physicians	51	82.3
Outpatient	43	69.4	Other	11	17.7
Inpatient	14	22.6	Medical specialty (could be >1)		
Other/unclear	5	8.1	GP/family physician	39	62.9
			Internists	25	40.3
			Other	20	32.3

can the effect of audit and feedback interventions be optimized?’

Based on our analyses, feedback seems most effective when it: is delivered by a supervisor or respected colleague; is presented frequently; includes both specific goals and action-plans; aims to decrease the targeted behavior; focuses on a problem where there was larger scope for improvement; and when the recipients are non-physicians. Unfortunately, relatively few trials feature these components. Furthermore, our findings suggest that investigators are not building upon best practices. For example, despite evidence that repeated feedback is more effective, studies that evaluate interventions after only one cycle of feedback continue to be performed. Furthermore, of the 32 studies conducted after 2002 considered in this analysis, feedback was delivered by a supervisor or respected colleague only six times, and no studies included feedback with both explicit goals and action plans. As a result, even after 140 randomized trials of audit and feedback, it remains difficult to identify how to optimize audit and feedback.<sup>6</sup> For instance, although a ‘supervisor or respected colleague’ appears to be the most effective source to deliver feedback, precise strategies to reliably identify and leverage such sources are not well known.<sup>13</sup> In addition, while it is advisable for action plans to accompany feedback since the

downside is minimal, the best way to operationalize this is unknown.<sup>7,14</sup> It is noteworthy that explicit targets without action plans do not seem to be particularly helpful. To achieve performance targets, recipients of feedback benefit from correct solution information<sup>8</sup> that can focus their attention on the targeted behavior(s).

Cumulative meta-analyses have previously been used to investigate whether future trials would be likely to change the conclusions regarding the effectiveness of QI or health services interventions.<sup>15,16</sup> For audit and feedback, it is plausible that further studies comparing the intervention against control may be informative if they are conducted for settings, professional groups or behaviors not well targeted in the current review (although relatively few additional trials should be needed to confirm whether observed effects are broadly aligned with observed effects across the body of literature). We recognize the risks of cumulative meta-analysis with respect to multiple testing and escalating type one error.<sup>17</sup> However, since the Cochrane review did not include a variance around the intervention effect, the figures showing the results of our cumulative analysis do not feature error bars as in the seminal examples of Lau et al.<sup>18</sup> Additionally, the number of characteristics tested in the meta-regression was limited by statistical and pragmatic concerns. Variables were only chosen for abstraction if there was an a priori directional



**Figure 3. Cumulative analysis—effect size\* of audit and feedback interventions over time (AF: audit and feedback; \*absolute difference in compliance with intended professional behaviors).**

hypothesis and a belief that data would be available in published reports. Confidence in the results of the meta-regression is limited by reliance upon indirect comparisons and risk of ecological fallacy. In other words, relationships identified across studies through meta-regression may not reflect relationships evident within studies; this is also known as aggregation bias. Finally, as with any review, the limitations of the primary studies must be considered.

We acknowledge that many other potential variables, including the clinical topic and context, likely impact the effectiveness of the intervention.<sup>19,20</sup> Amongst the 98 comparisons, there were 41 comparisons testing audit and feedback alone and 57 comparisons testing audit and feedback as the core, essential part of a multifaceted intervention. It is plausible that co-interventions may interact with the effect modifiers tested in the meta-regressions. A recent international meeting was conducted to identify high-yield research questions for understanding how to enhance the effectiveness of audit and feedback. Stakeholders suggested a need for more research to better understand how contextual and recipient characteristics

moderate audit and feedback effectiveness, characteristics of the desired behavior change that make a good target for audit and feedback, and how the specific design of the audit and feedback intervention interacts with these factors.<sup>21</sup>

Given the importance of audit and feedback as a key component of many QI interventions, there is a need to identify opportunities to sequentially and systematically test various approaches to the design and development of audit and feedback. Researchers can continue to conduct uncoordinated trials of audit and feedback versus usual care and rely upon periodically conducted meta-regressions across studies to explore effect modifiers. But the results will be at risk of ecological fallacies, and as demonstrated here, this approach has resulted in minimal advances over time. Alternatively, researchers could achieve greater confidence in causal inference regarding more effective intervention design through a limited number of multi-arm trials with direct, head-to-head comparisons testing different approaches for designing and delivering audit and feedback. Another approach that could help advance cumulative knowledge regarding audit and feedback and other QI

Table 3. Factors Explaining Variability in Effectiveness of Feedback: Serial Meta-Regressions

Characteristic of feedback	Estimated effect size*, (no. studies)		
	2010	2006	2002
Format of feedback	$p=0.386$	$p=0.731$	$p=0.729$
Verbal	12.77, (15)	14.85, (14)	17.02, (12)
Written	20.70, (50)	19.94, (41)	23.76, (19)
Both verbal and written	19.05, (27)	19.19, (26)	16.98, (18)
Not clear	16.90, (6)	13.58, (5)	2.94, (2)
Source of feedback	$p=0.006$	$p=0.034$	$p=0.300$
A supervisor or respected colleague	25.22, (10)	23.49, (8)	24.48, (4)
Standards review org. or representative of employer	9.16, (3)	9.38, (3)	0.90, (1)
The investigators	15.19, (52)	14.71, (42)	17.85, (13)
Not clear	19.85, (33)	19.99, (33)	17.47, (33)
Frequency of feedback	$p<0.001$	$p<0.001$	$p<0.001$
Frequent (up to weekly)	27.58, (5)	28.50, (3)	28.64, (2)
Moderate (up to monthly)	18.51, (10)	16.73, (9)	18.31, (4)
Infrequent (less than monthly)	14.04, (26)	13.32, (22)	1.06, (10)
Once only	7.49, (52)	7.75, (47)	9.96, (30)
Unclear;	19.15, (5)	18.17, (5)	17.92, (5)
Instructions for improvement	$p=0.044$	$p=0.068$	$p=0.325$
Explicit, measurable target, but no action plan	10.88, (5)	10.45, (5)	8.48, (1)
Action plan, but no explicit target	17.16, (32)	16.69, (31)	11.37, (18)
Both	23.19, (4)	23.06, (4)	22.01, (4)
Neither;	18.18, (57)	17.37, (46)	18.84, (28)
Nature of change required	$p=0.025$	$p=0.028$	$p=0.510$
Increase current behavior	15.55, (40)	15.65, (36)	19.34, (17)
Decrease current behavior	22.46, (11)	22.30, (11)	12.61, (4)
Change behavior to similar alternative or unclear	14.05, (47)	12.73, (39)	13.58, (30)
Profession of recipient (Physician yes/no)	$p<0.001$	$p<0.001$	$p<0.001$
Physician	10.99, (82)	10.19, (72)	4.80, (45)
Not physician	23.72, (16)	23.60, (14)	25.55, (6)
Risk of bias	$p=0.375$	$p=0.564$	$p=0.281$
Yes (low risk of bias)	14.85, (32)	14.92, (27)	21.34, (8)
Unclear	15.79, (51)	15.33, (48)	10.06, (34)
No (high risk of bias);	21.42, (15)	20.43, (11)	14.12, (9)
Baseline performance (continuous variable)	$p<0.001$	$p=0.003$	$p=0.021$

\*Absolute difference in compliance with intended professional behaviors

strategies would be to consider engineering-based methodological options that enable testing of multiple potential effect modifiers, such as theory-driven factorial and/or sequential adaptive trials.<sup>22</sup> Future audit and feedback interventions should feature the aspects known to be associated with greater effectiveness and future trials should be powered to find relatively small effect sizes, especially in the case of head-to-head trials. This proposed shift in direction for QI trials parallels the movement to limit placebo-controlled trials of clinical interventions and to increase focus on comparative effectiveness research.<sup>23</sup>

The findings of this review suggest that QI trialists have failed to cumulatively learn from previous studies (or from systematic reviews). Rather, it would appear that the norm for those testing audit and feedback interventions is to 're-invent the wheel', repeating rather than learning from and contributing to extant knowledge.<sup>24</sup> As highlighted in the recent series on increasing value and reducing waste in research,<sup>25</sup> the opportunity cost of continuing in the current manner is large for patients, providers, and health systems. A coordinated approach toward building upon previous literature and relevant theory to identify the key, active ingredients of interventions would help QI stakeholders achieve greater impact with their interventions and produce outcomes

that are more generalizable.<sup>26,27</sup> In particular, QI trialists could benefit from adapting the model of the Children's Oncology Group, which has successfully shared resources to accelerate progress.<sup>28</sup> At a minimum, for stakeholders involved in the funding and conduct of QI trials, this analysis emphasizes the need for trials of carefully planned interventions with explicitly justified components to ensure that the field of QI in healthcare can move forward.

**Acknowledgements:** This study was conceived by NMI and JMG. Analyses were conducted by JOJ. The first draft was written by NMI and revised with critical input from all authors. All authors approved the final version of the manuscript. This study received no specific external funding. NMI is supported by research fellowships from the Canadian Institutes of Health Research and the Department of Family and Community Medicine, University of Toronto. JMG is supported by a Canada Research Chair in Knowledge Transfer and Uptake.

**Conflict of Interest:** The authors declare that they do not have a conflict of interest.

**Corresponding Author:** Noah M. Ivers, MD, PhD; Family Practice Health Centre and Institute for Health Systems Solutions and Virtual Care, Women's College Hospital, Toronto, Ontario, Canada (e-mail: noah.ivers@utoronto.ca).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## REFERENCES

- Davis DA, Mazmanian PE, Fordis M, Van Harrison RTKE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA*. 2006;296(9):1094.
- Hysong SJ, Teal CR, Khan MJ, Haidet P. Improving quality of care through audit and feedback. *Implement Sci*. 2012;7:45. doi:10.1186/1748-5908-7-45.
- Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev*. 2012;6, CD000259.
- Jamtvedt G, Young JM, Kristoffersen DT, Thomson O'Brien MA, Oxman AD. Audit and feedback: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev*. 2003;3(3), CD000259.
- Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD. Does telling people what they have been doing change what they do? A systematic review of the effects of audit and feedback. *Qual Saf Health Care*. 2006;15(6):433-6.
- Foy R, Eccles MP, Jamtvedt G, Young J, Grimshaw JM, Baker R. What do we know about how to do audit and feedback? Pitfalls in applying evidence from a systematic review. *BMC Health Serv Res*. 2005;5:50.
- Gardner B, Whittington C, Mcateer J, Eccles MP, Michie S. Using theory to synthesise evidence from behaviour change interventions: the example of audit and feedback. *Soc Sci Med*. 2010;70(10):1618-25.
- Hysong SJ. Meta-analysis: audit and feedback features impact effectiveness on care quality. *Med Care*. 2009;47(3):356-63.
- Colquhoun HL, Brehaut JC, Sales A, Ivers N, Grimshaw J, Michie S, et al. A systematic review of the use of theory in randomized controlled trials of audit and feedback. *Implement Sci*. 2013;8(1):66.
- Suggested Risk Of Bias Criteria For EPOC Review. Cochrane Effective Practice And Organisation Of Care (EPOC) Group. 2013. Available At: <http://epoc.cochrane.org/sites/epoc.cochrane.org/files/uploads/14%20Suggested%20risk%20of%20bias%20criteria%20for%20EPOC%20reviews%202013%2008%2012.Pdf> (Accessed 2 May 2014).
- Higgins JPT, Green S. Cochrane Handbook For Systematic Reviews Of Interventions. Cochrane Database Of Systematic Reviews. 2008. Available at: <http://handbook.cochrane.org> (Accessed 2 May 2014).
- Grimshaw JM, Thomas RE, Maclennan G, Fraser C, Ramsay CR, Vale L, et al. Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technol Assess*. 2004;8(6):1-72.
- Flodgren G, Parmelli E, Doumit G, Gattellari M, O'Brien MA, Grimshaw J. Local opinion leaders: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev*. 2011;8(8), CD000125.
- Ivers NM, Tu K, Francis J, Barnsley J, Shah B, Upshur R, et al. Feedback GAP: study protocol for a cluster-randomized trial of goal setting and action plans to increase the effectiveness of audit and feedback interventions in primary care. *Implement Sci*. 2010;5:98.
- Gilbody S, Bower P, Fletcher J, Richards D, Sutton AJ. Collaborative care for depression: a cumulative meta-analysis and review of longer-term outcomes. *Arch Intern Med*. 2006;166(21):2314-21.
- Dent L, Taylor R, Jolly K, Raftery J. "Flogging dead horses" evaluating when have clinical trials achieved sufficiency and stability? A case study in cardiac rehabilitation. *Trials*. 2011;12:83. doi:10.1186/1745-6215-12-83.
- Pogue JM, Yusuf S. Cumulating evidence from randomized trials: utilizing sequential monitoring boundaries for cumulative meta-analysis. *Control Clin Trials*. 1997;18(6):580-93. Discussion 661-6.
- Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N Engl J Med*. 1992;327(4):248-54.
- Van Der Veer SN, De Keizer NF, Ravelli AC, Tenkink S, Jager KJ. Improving quality of care. A systematic review on how medical registries provide information feedback to health care providers. *Int J Med Inform*. 2010;79(5):305-23.
- Hysong SJ, Best RG, Pugh JA. Audit and feedback and clinical practice guideline adherence: making feedback actionable. *Implement Sci*. 2006;1:9.
- Ivers NM, Sales A, Colquhoun HL, Michie S, Foy R, Francis JJ, et al. No more 'business as usual' with audit and feedback interventions: towards an agenda for a reinvigorated intervention. *Implement Sci*. 2014;9:14.
- Collins LM, Murphy SA, Nair V, Strecher VJ. A strategy for optimizing and evaluating behavioural interventions. *Ann Behav Med*. 2005;30(1):65-73.
- Chokshi DA, Avorn J, Kesselheim AS. Designing comparative effectiveness research on prescription drugs: lessons from the clinical trial literature. *Health Aff (Millwood)*. 2010;29(10):1842-8.
- Auerbach AD, Landefeld CS, Shojania KG. The tension between needing to improve care and knowing how to do it. *N Engl J Med*. 2007;357(6):608-13.
- Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ionnadis JP, et al. Biomedical value: increasing value, reducing waste. *Lancet*. 2014;383(9912):101-4.
- Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ*. 2008;337:A1655.
- Goudie AC, Sutton AJ, Jones DR, Donald A. Empirical assessment suggests that existing evidence could be used more fully in designing randomized controlled trials. *J Clin Epidemiol*. 2010;63(9):983-91.
- O'Leary M, Krailo M, Anderson JR, Reaman GH, Children's Oncology Group. Progress in childhood cancer: 50 years of research collaboration, a report from the Children's Oncology Group. *Semin Oncol*. 2008;35(5):484-93.