

# The Role of Peers in General Practice Quality Improvement: An Overview of Systematic Reviews\*

[Click here for the latest version.](#)

Theodore Caputi

March 23, 2020

## 1 Purpose

The purpose of this overview is to provide a qualitative synthesis of the evidence of efficacy for different types of peer-based interventions targeting health care quality improvement. There are several types of peer-based interventions targeting health care quality that may act through similar behavior change channels, and this Overview is intended to merge the research from these different types of interventions. The research question, then, is: “What evidence of efficacy exists for peer-based interventions to improve healthcare quality in the General Practice setting?”.

## 2 Methods

The Methods of this overview are intended to follow Cochrane Symposium guidance for Overviews of Reviews ([Hunt et al., 2018](#)), with minimal semantic differences to improve clarity.

---

\*These results are preliminary.

## 2.1 Criteria for Considering Reviews for Inclusion

### 2.1.1 Types of Interventions

Reviews published in peer-reviewed journals that describe the efficacy of a peer-based intervention(s) to improve health care quality will be included, so long as the intervention(s) of focus meet the following criteria:

- The stated purpose of the intervention or outcome of interest is to improve at least one structure, process, or clinical outcome indicator or a class of indicators.
- The intervention targets physicians, with a preference for interventions targetting general practitioners.
- The predominant vehicle of the intervention is through peer interaction.
- The intervention should be personal in nature (e.g., not mass messages sent from a governing body that happens to include general practice physicians as signatories), preferably including face-to-face contact.

Systematic reviews of quality improvement interventions can generally be divided into those that focus on the effect of a particular type of intervention on general quality (one intervention on many outcomes; e.g., “Paying for performance to improve the delivery of health interventions in low- and middle- income countries”), or the set of interventions used to improve a specific metric of quality (many interventions on one outcome; e.g., “Interventions to optimise prescribing for older people in care homes”). Some other systematic reviews focus on the evidence of one type of intervention on one outcome. Because the purpose of this Overview is to synthesize evidence of the efficacy for different types of peer-based intervention, only reviews that focus on peer-based interventions are included. That is, reviews that, on the whole, do not aim to assess the efficacy of a specific type of peer-based intervention will be excluded, even if some of the primary studies in the review focus on peer-based interventions.

### 2.1.2 Types of Designs

Peer-based interventions can be studied using several experimental and quasi-experimental methods, and it is likely that a given review will include primary studies with a mix of methodologies. In these cases, the Overview will “include” the entire review but focus on results from the following study designs: randomised controlled trial; non-randomised trial; difference-in-difference studies; interrupted time series analysis (ITSA), including both one-sample ITSA and two-sample ITSA. Because these study designs vary in their robustness, different inclusion criteria were applied to each (see Table 1). Importantly, both experimental studies (where the investigators impose an intervention they designed) and quasi-experimental studies (where the investigators use observational data surrounding an event) will be considered.

## 2.2 Types of Participants

Reviews of interventions that target general practitioners or specialist physicians will be included in the Overview. Those that target other healthcare providers (e.g., nurses) will be excluded.

The primary studies within the reviews may vary in their level of analysis. Some primary studies (e.g., those that focus on structure and process outcomes) may use physician, practice, ward, or region-level outcomes, while other primary studies (e.g., those that focus on clinical outcomes) may use patient-level outcomes. Both types of primary studies will be considered in the Overview.

## 2.3 Types of Outcomes

Reviews that focus on the intervention’s effects on overall quality or on changes in structure, process, or clinical outcomes are considered. Ideally, studies should choose these measures *a priori* to their analyses, limiting the risk of investigator-induced bias ([Ioannidis, 2005](#)).

Table 1: Study Designs and Inclusion Criteria

Study Design	Inclusion Criteria
Randomised controlled trial	<ul style="list-style-type: none"> <li>• True, reliable randomisation procedure at the individual or block level</li> <li>• If block-randomised, at least four blocks</li> </ul>
Non-randomised trial	<ul style="list-style-type: none"> <li>• Authors provide adequate explanation for why trial could not feasibly be randomised (i.e., decision driven by circumstance rather than investigator choice)</li> <li>• Contains control group</li> <li>• If trial is organised by blocks, at least two control groups and two intervention groups</li> </ul>
Two-sample interrupted time series <sup>1</sup>	<ul style="list-style-type: none"> <li>• Adequately robust description of parallel pre-trends between the control and treatment group</li> <li>• Preferably using longitudinal data but, if not, it is reasonable to assume that the underlying samples are the same at different time points</li> <li>• Control group adequately similar to treatment group</li> <li>• Adjusted for relevant confounders</li> <li>• No known external event occurring at around the same time as the intervention</li> </ul>
One-sample Interrupted time series	<ul style="list-style-type: none"> <li>• Adequately precise, linear pre-trend</li> <li>• Preferably using longitudinal data but, if not, it is reasonable to assume that the underlying samples are the same at different time points</li> <li>• Separate estimates for immediate and slope effect or mean overall effect</li> <li>• No known external event occurring at around the same time as the intervention</li> <li>• Adjusted for relevant confounders</li> </ul>
Difference-in-difference <sup>1</sup>	<ul style="list-style-type: none"> <li>• Groups adequately similar at baseline</li> <li>• Preferably using longitudinal data but, if not, it is reasonable to assume that the underlying samples are the same at different time points</li> <li>• No external event occurring between pre- and post-period</li> <li>• Adjusted for relevant confounders</li> </ul>

**NOTE:** These design-specific inclusion criteria were chosen based upon the assumptions of the underlying models. Adequacy will be determined by the author, who will err on the side of caution in including studies.

<sup>1</sup> There is no restriction on the number of pre- or post-periods for two-sample ITSA because a two-period, two-sample ITSA is equivalent to a difference-in-difference design study.

Main outcomes from these systematic reviews may include the following:

- Physician professional behavior (e.g., adherence to clinical guidelines, referral rates, or prescribing rates)
- Healthcare organisation efficiency and efficacy, such as waiting list time or achievement on preset quality measures
- Patient-level outcomes, such as patient morbidity and/or mortality and patient satisfaction

Secondary outcomes from these systematic reviews may include the following:

- Adverse events
- Economic and cost-effectiveness analyses

## **2.4 Other Inclusion/Exclusion Criteria**

Some additional inclusion and exclusion criteria were applied to the studies.

### **Inclusion Criteria**

- Published in peer-reviewed journal
- Systematic review (including systematically performed scoping reviews but excluding non-systematic or narrative reviews)
- When systematic reviews have been updated over time, all previous iterations will be considered (though the latest review will be noted)

### **Exclusion Criteria**

- Systematic reviews with inadequate search strategies, defined as searching less than two DARE sources
- Studies not published in English (due to limitations of the Overview author)

Table 2: Terms for Search Strategy

<b>List A: Quality Improvement</b> 79,711 Matching Articles	<b>List B: Physician Targets</b> 773,406 Matching Articles
“quality improvement” “healthcare quality” “health care quality” “professional practice”	“physician(s)” “clinician(s)” “doctor(s)”
<b>List C: Peer-Based</b> 450,761 Matching Articles	
“peer” “opinion leader” “influential physician” “educationally influential” “social network” “audit” “feedback”	

**NOTE:** These search terms were selected based upon related past reviews and iterative searching by the author on PubMed. Quotes indicate that some portion of the article must match the entire term. Parentheses with an asterisk indicate that the term may take on a suffix in the article. For example, the search term (audit\*) would include all articles that use the terms “audits”, “auditors”, and “auditing”. Article counts by list are accurate for PubMed as of 2 December 2019.

## 2.5 Search Methods

I systematically search databases for published, peer-reviewed systematic reviews that assess the efficacy of peer-based interventions in improving healthcare quality. Specifically, I screened all studies resulting from a search that contains at least one term from list A, at least one term from list B, and at least one term from list C (Table 2). This search strategy was initially implemented in the PubMed database, using the RISmed package for R version 3.6.1.

## 2.6 Searching Other Resources

After the initial selection of reviews, studies in the bibliographies of initially included articles were also considered for inclusion.

## 3 Data Extraction and Analysis

### 3.1 Selection of Studies

For each review emerging from the initial search, the title, year, journal, and abstract is exported to a spreadsheet in Microsoft Excel. The first screen uses only the review title and the abstract. If, based upon the review’s title and/or abstract, it is clear the review fit the inclusion criteria, then the review is coded as “Eligible”. If, based upon the title and/or abstract, the review is clearly outside of the scope of the overview, it is coded as “Ineligible”. Remaining reviews are coded as “Unclear if Eligible”.

The full text of reviews coded as either “Eligible” or “Unclear if Eligible” are downloaded and reviewed. The full text is used to make a final decision on whether the review meets or fails to meet the inclusion/exclusion criteria from Section 2.1. Those that meet inclusion/exclusion criteria are included in the review. Those that fail to meet the inclusion/exclusion criteria are coded for which criteria they most clearly failed to meet. After the first round of reviews are included, those initial reviews’ bibliographies are screened for other relevant reviews. In cases where the review’s full text could not be downloaded (via University of York or partner university databases), the omission is noted.

The results of this process are reported in a PRISMA diagram.

### 3.2 Data Extraction and Management

A data collection tool was created based upon ([Flodgren et al., 2011](#)). This data extraction tool was piloted on one inclusion criteria and then modified to capture relevant information.

Ultimately, details of the stated aim, search strategy, number of included primary studies, number of participants, primary study designs, primary study intervention type, primary study comparison group, primary study country of focus, primary study healthcare setting focus, and reviewer interpretation were captured for each included review. When available, information on the possibility of publication bias (e.g., Egger’s Test ([Sterne and Egger, 2001](#))) is also recorded.

### **3.3 Quality of Included Reviews**

Along with the information collected for each review in the data collection tool, the quality of each included review will be assessed through the AMSTAR 2, a validated critical appraisal tool intended to evaluate the quality of systematic reviews that, themselves, include randomised and non-randomised studies of healthcare interventions ([Shea et al., 2017](#)).

### **3.4 Assessment of Risk of Bias in Included Studies**

Quality of primary studies is reported from each systematic review’s quality assessment. In cases where a review’s quality assessment is inadequate, quality of the underlying primary studies is assessed in terms of the inclusion criteria presented in Table 1. For example, a one-sample ITSA is assessed for adequately precise, linear pre-trends; measuring the same sample in different time points; adjustment for relevant confounders; and presence of an external event.

### **3.5 Synthesis**

Because the aims of included reviews are highly varied, a quantitative meta-analysis would be largely uninformative. For example, reviews may seek to assess the efficacy of a peer-based intervention in terms of improving patient waiting times while another assesses the efficacy of a peer-based intervention in terms of adherence to clinical guidelines for treating



cardiovascular problems. These two outcomes cannot reasonably be compared, and neither should the reviews. Instead, the papers will be summarised, categorised, and synthesised. The qualitative review attempts to (A) document which types of peer-based interventions have been assessed in systematic review, (B) describe the available evidence of efficacy for each of these types of peer-based interventions, and (C) highlight common threads emerging from the research on several types of peer-based interventions.

If a review provides some quantitative estimate of effect, this will be included, in its original form, along with the review's summary. However, effect sizes will not be directly compared against one another. If possible, when a review does not contain a quantitative effect estimate, the direction of the effect (e.g., improves quality, does not improve quality, mixed results) will be reported.

### **3.6 Limitations**

This design for an Overview of systematic reviews has several limitations. Only indexed, peer-reviewed literature is considered in the Overview. While this provides a useful quality screen for included reviews, it is possible that high-quality literature that would otherwise fit the inclusion criteria are available. Due to the constraints of the author, only reviews written in English are included. Consequently, there may be relevant reviews published in other languages. Because the review focuses different types of intervention that could each be applied to several outcomes rather than a specific intervention and specific outcome, results are not quantitatively comparable. This presents several limitations. For example, it is not possible to estimate the average effect of peer-based treatments, generally. Further, sensitivity or subgroup analyses cannot be conducted.

## 4 Results

Four major types of peer-based interventions have been systematically reviewed in the literature: feedback, education, opinion leaders, and governance.

### 4.1 Feedback

The plurality ( $n=6$ ) of the reviews focused on feedback-based interventions. Feedback is a mechanism whereby an individual is told how his or her practice measures against some standard or target ([Ivers et al., 2012](#)). Feedback interventions are often necessarily paired with audit, which is the preliminary step where data on an individual's performance is collected and compared to a standard. Not all feedback mechanisms fit within the peer framework. For example, an audit and feedback mechanism may simply compare a GP's performance to a standard set by a governing committee. However, audit and feedback often involve peers; for example, a peer-based feedback mechanism may involve having a peer physician discuss the individual's audit results.

The most current systematic review of feedback interventions was conducted as a full Cochrane Review by [Ivers et al. \(2012\)](#). Of the 140 studies included in the review, the person who provided the feedback was either unclear or the investigators in 112. Of the remaining 28, 13 used feedback from a supervisor or senior colleague, and 15 used feedback from an impersonal professional standards organisation or representative of the employers. The authors conducted a meta-regression over all studies without high risk of bias to compute the expected adjusted risk difference (RD) between those that received feedback from different sources. The adjusted RD for feedback provided by a supervisor or senior colleague was 16.50, compared with 2.37 for a professional standards review organisation or the employer, 5.04 for the investigators, and 5.48 for unclear source of feedback ( $p<0.001$ ), suggesting that receiving feedback from peers provided a significant advantage over receiving feedback from other sources. The authors conclude that feedback mechanisms are most successful when “the

person responsible for the audit and feedback is a supervisor or colleague” (Ivers et al., 2012), indicating that peer influence plays a key role in the effectiveness of feedback mechanisms. The same research team reached a similar conclusion in another analysis using the data from the Cochrane Review; Ivers et al. (2014) found in a median-of-medians meta-regression (which included those primary studies with high risk of bias) that interventions that use feedback from a supervisor or respected colleague had a significantly greater effect than interventions that used feedback from other sources. The estimated adjusted RD among the 10 studies with feedback from a supervisor or respected colleague was 25.22, compared with 9.16 among the 3 studies using feedback from standards review organisations or employers, 15.19 among the 52 studies using feedback from the investigators, and 19.85 among the 33 studies where the source of feedback was not clear ( $p=0.006$ ).

Importantly, Ivers et al. (2012) did not report the quality of primary studies individually, and so it is not possible to evaluate whether there was a difference in quality of the primary studies that used peer-based feedback versus other forms of feedback. However, this information is available in past reviews of the same subject. In a preceding Cochrane review from 2003 (Jamtvedt et al., 2003), there were also 13 studies that used feedback from supervisors or senior colleagues; 3 (23%) of these were assessed to have low quality, 6 (46%) were assessed to have moderate quality, and 4 (31%) to have high quality. Compared with other studies in the same review, those peer-based feedback appear to have higher quality than those without peer-based feedback (31% high quality versus 8% high quality).

In the 2006 iteration of the Cochrane review, Jamtvedt et al. (2006b) describe two specific studies that deal with feedback from peers (Jamtvedt et al., 2006a). The first study (Homborgh et al., 1999) used an RCT design to randomise 90 physicians in 68 practices to receive audit and feedback with a visit from a physician or audit and feedback with a visit from a non-physician observer. The authors found that, while both interventions were effective, the impact on performance relative to 208 performance indicators was better among the treatment group who had received outreach from a physician than those who had received

outreach from a non-physician observer. The second study ([Ward et al., 1996](#)) compared the effect on diabetes management of audit and feedback with outreach from a physician versus audit and feedback with outreach from a nurse. The authors did not find a significant difference between the effect size of each treatment, as assessed through the Adequate Competent Care Score.

A separate systematic review ([Ferguson, Wakeling and Bowie, 2014](#)) focused on feedback emerging from colleagues (including other physicians and non-medical co-workers), which the authors refer to as “Multisource Feedback”, in preparation for the introduction of multisource feedback into the UK General Medical Council scheme for reviewing all medical doctors. The authors ultimately reviewed 16 studies, including 1 RCT, 7 cross-sectional survey studies, 2 mixed methods studies, and 6 qualitative studies. According to the reviewers, only the RCT identified a “measured change in behaviour” emerging from multisource feedback. Those physicians in the treatment group received a tailored coaching session and multisource feedback. Compared with the control group, the treatment group had a 35% (95%CI 11.0-58.0) increase in communicating effectively with the patient and family, a 30% (95%CI 7.9-53.0) increase in timeliness of completing tasks, and a 26% (95%CI 2.9-49.0) increase in demonstrating responsibility and accountability. The reviewers point out that it is not possible to disentangle the effect of the feedback from the tailored coaching session, as the control group did not receive the tailored coaching session, and that the sample size was small ( $n=36$ ). Twelve of the remaining primary studies reviewed assessed changes in behavior through self-report. According to the reviewers, the most commonly reported change was an improvement in communication, with some studies also reporting improvements in patient follow-up, medical record maintenance, providing information to patients, and managing stress.

Within this review ([Ferguson, Wakeling and Bowie, 2014](#)), three primary studies investigated the characteristics of the raters on the impact of the feedback. According to the reviewers, these primary studies collectively suggested that feedback was viewed as more ac-

ceptable and more effective when the recipient viewed the rater as familiar with the work. The review contained six primary studies in which feedback was facilitated by a colleague (i.e., appraiser, mentor, facilitator, supervisor, or coach) and eleven primary studies where feedback was not facilitated (i.e., provided in paper report). While a direct comparison between the forms of facilitation is not made, the reviewers state that facilitation was key to the physicians' acceptance of and response to the multisource feedback.

Finally, one systematic review examined the effectiveness of audit and feedback mechanisms in the context of a single disease. [Guldborg et al. \(2009\)](#) conducted a systematic review of randomised controlled trials focused on the effect of feedback on quality of care for people with Type II diabetes in the general practice setting. Of the ten primary studies that met their inclusion criteria, two used an intervention that involved a significant peer component. The first study ([Phillips et al., 2005](#)) randomised clinicians to either receive feedback sessions with an endocrinologist, receive computerised reminders, or both, but found that only the combination of the two interventions improved patient-level clinical outcomes (assessed through mean HbA1c). The second study RCT ([Frijling et al., 2002](#)) randomised clinicians to receive feedback reports and outreach visits, finding mixed results; the intervention improved outcomes on two of seven dimensions (percentage of recommended foot examinations and eye examinations performed).

## 4.2 Local Opinion Leaders

One Cochrane Review, updated four times since 1999, concerns the effectiveness of local opinion leaders on healthcare quality. The most current iteration of this review ([Flodgren et al., 2019](#)) included 24 randomised controlled trials and assessed the effectiveness of opinion leader based interventions on several types of quality improvement outcomes. For example, in a meta-analysis of 71 dichotomous outcomes relating to adherence to evidence-based practice across 18 studies, the authors found moderate evidence that opinion leaders “probably improve healthcare professionals’ compliance with evidence-based practice”, with a median

adjusted risk difference of 10.8% (IQR: 3.5% to 14.6%). On the other hand, using the results from 8 primary studies, the reviewers found that the evidence that opinion leader based interventions improved patient outcomes was mixed. Some interventions found improvements in patient outcomes, such as a 45% (95%CI 9-71) decrease in the rate of women experiencing a  $\geq 500$  mL reduction in postpartum haemorrhage and a 0.25 mean decrease ( $p < 0.05$ ) in parent-reported emergency department visits for children with asthma. However, opinion leader based interventions failed to have an impact on several indicators within these primary studies; for example, opinion leader interventions failed to significantly affect pain scores for people with cancer (mean difference +0.86 scale steps;  $p = 0.66$ ), to affect local cancer recurrence (risk difference +0.1%) in patients with colon cancer, or improve functional rating scores for people who had experienced whiplash (median difference -0.6, 95%CI -7.8 to 6.6). Ultimately, the reviewers concluded that the certainty of evidence that opinion leader based interventions improve patient outcomes is very low. Among the 24 primary studies, the reviewers found that 11 were at low risk of bias, while 13 were at moderate or high risk of bias. However, they also found that the median effect size did not meaningfully vary between those studies at low risk for bias and those at moderate or high risk for bias.

Several of the primary studies in the review investigated interventions that combined the opinion leaders with other interventions, and so the reviewers also decomposed their analysis based upon the treatment intervention (e.g., opinion leader alone or opinion leader with another intervention) and the comparison intervention (e.g., no intervention, another intervention, or the same other intervention as the treatment group). Among the 38 outcomes across five primary studies with the cleanest comparison, i.e., opinion leader alone versus no intervention, the authors found moderate-certainty evidence that opinion leader interventions “probably improve healthcare professionals’ compliance with evidence-based practice”. The median adjusted risk difference for these interventions was 9.15% (95%CI -0.3% to 15%).

### 4.3 Education

Two included reviews ([O'Brien et al., 1997](#), [2007](#)) focused on educational interventions. The most current iteration ([O'Brien et al., 2007](#)) is a Cochrane Review of RCTs that investigate the effect of Educational Outreach Visits (EOV), which the authors define as a face-to-face visit by an external trained healthcare professional that occurs at the target physician's workplace and is intended to improve quality of care. According to the authors, EOVs incorporate interventions that had previously been referred to as educational detailing, public interest detailing, academic detailing, and educational visiting. Reviewing 69 trials attempting to address prescribing behavior (n=29), increase preventive screening (n=11), and other problems in general practice (n=29), the authors found that EOVs have a small and variable effects on healthcare quality. For the 49 trials using dichotomous outcomes, the median adjusted risk difference for EOV interventions was 5.6% (IQR: 3% to 9%), and quality improvement exceeded more than 20% in half of the 20 trials that used continuous outcomes. Still, the authors note that quality improvements from EOVs were often well below the primary study authors' expectations. While the authors did not decompose these effects based upon what type of person delivered the EOV, most interventions included in the review (n=53) used physicians and could, consequently, be considered peer-based interventions.

Just one primary study in the review (which was also included in the Cochrane Review on audit and feedback ([Jamtvedt et al., 2006b](#))) specifically studied whether EOVs (combined with an audit and feedback intervention) were more effective in improving practice management when the presenter was a physician relative to a trained non-physician. As previously mentioned, [Homborgh et al. \(1999\)](#) found that, while both groups improved practice management, those with peer EOV presenters had greater improvements; those in the peer EOV group improved in dimensions of collaboration and practice organisation, while those in the non-physician EOV group improved in two outcomes related to maintaining patient records.

## 5 Discussion

The results of the Overview suggest that the effectiveness of peer-based interventions for quality improvement in healthcare are well documented. There is substantial evidence that peer-based feedback, local opinion leaders, and education interventions can improve health care quality. However, only these three types of peer-based interventions have been thoroughly investigated in systematic reviews. Other peer-based interventions



## References

- Ferguson, Julie, Judy Wakeling, and Paul Bowie.** 2014. “Factors influencing the effectiveness of multisource feedback in improving the professional practice of medical doctors: a systematic review.” *BMC Medical Education*, 14(1): 76.
- Flodgren, Gerd, Martin P Eccles, Sasha Shepperd, Anthony Scott, Elena Parmelli, and Fiona R Beyer.** 2011. “An overview of reviews evaluating the effectiveness of financial incentives in changing healthcare professional behaviours and patient outcomes.” *Cochrane Database of Systematic Reviews*.
- Flodgren, Gerd, Mary Ann O’Brien, Elena Parmelli, and Jeremy M. Grimshaw.** 2019. “Local opinion leaders: effects on professional practice and healthcare outcomes.” *Cochrane Database of Systematic Reviews*, , (6).
- Frijling, B. D., C. M. Lobo, M. E. J. L. Hulscher, R. P. Akkermans, J. C. C. Braspenning, A. Prins, J. C. van der Wouden, and R. P. T. M. Grol.** 2002. “Multifaceted support to improve clinical decision making in diabetes care: a randomized controlled trial in general practice.” *Diabetic Medicine: A Journal of the British Diabetic Association*, 19(10): 836–842.
- Guldborg, Trine Lignell, Torsten Lauritzen, Jette Kolding Kristensen, and Peter Vedsted.** 2009. “The effect of feedback to general practitioners on quality of care for people with type 2 diabetes. A systematic review of the literature.” *BMC family practice*, 10: 30.
- Hombergh, P. van den, R. Grol, H. J. van den Hoogen, and W. J. van den Bosch.** 1999. “Practice visits as a tool in quality improvement: mutual visits and feedback by peers compared with visits and feedback by non-physician observers.” *BMJ Quality & Safety*, 8(3): 161–166.

- Hunt, Harriet, Alex Pollock, Pauline Campbell, Lise Estcourt, and Ginny Brunton.** 2018. “An introduction to overviews of reviews: planning a relevant research question and objective for an overview.” *Systematic Reviews*, 7(1): 39.
- Ioannidis, John P. A.** 2005. “Why Most Published Research Findings Are False.” *PLOS Medicine*, 2(8): e124.
- Ivers, Noah, Gro Jamtvedt, Signe Flottorp, Jane M Young, Jan Odgaard-Jensen, Simon D French, Mary Ann O’Brien, Marit Johansen, Jeremy Grimshaw, and Andrew D Oxman.** 2012. “Audit and feedback: effects on professional practice and healthcare outcomes.” *Cochrane Database of Systematic Reviews*.
- Ivers, Noah M., Jeremy M. Grimshaw, Gro Jamtvedt, Signe Flottorp, Mary Ann O’Brien, Simon D. French, Jane Young, and Jan Odgaard-Jensen.** 2014. “Growing Literature, Stagnant Science? Systematic Review, Meta-Regression and Cumulative Analysis of Audit and Feedback Interventions in Health Care.” *Journal of General Internal Medicine*, 29(11): 1534–1541.
- Jamtvedt, G., J. M Young, D. T Kristoffersen, M. A. O’Brien, and A. D Oxman.** 2006a. “Does telling people what they have been doing change what they do? A systematic review of the effects of audit and feedback.” *Quality and Safety in Health Care*, 15(6): 433–436.
- Jamtvedt, G, Jm Young, Dt Kristoffersen, Ma Thomson O’Brien, and Ad Oxman.** 2003. “Audit and feedback: effects on professional practice and health care outcomes.” In *The Cochrane Database of Systematic Reviews*. , ed. The Cochrane Collaboration, CD000259. Chichester, UK:John Wiley & Sons, Ltd.
- Jamtvedt, Gro, Jane M Young, Doris Tove Kristoffersen, Mary Ann O’Brien, and Andrew D Oxman.** 2006b. “Audit and feedback: effects on professional practice and

health care outcomes.” In *Cochrane Database of Systematic Reviews*. , ed. The Cochrane Collaboration, CD000259.pub2. Chichester, UK:John Wiley & Sons, Ltd.

**O’Brien, Mary Ann, Andy D. Oxman, Dave Davis, R. B. Haynes, Nick Freemantle, and Emma Harvey.** 1997. “Educational outreach visits: effects on professional practice and health care outcomes.” *Cochrane Database of Systematic Reviews*, , (4).

**O’Brien, Mary Ann, Stephen Rogers, Gro Jamtvedt, Andrew D Oxman, Jan Odgaard-Jensen, Doris Tove Kristoffersen, Louise Forsetlund, Daryl Bainbridge, Nick Freemantle, Dave Davis, R. Brian Haynes, and Emma Harvey.** 2007. “Educational outreach visits: effects on professional practice and health care outcomes.” *Cochrane Database of Systematic Reviews*.

**Phillips, Lawrence S., David C. Ziemer, Joyce P. Doyle, Catherine S. Barnes, Paul Kolm, William T. Branch, Jane M. Caudle, Curtiss B. Cook, Virginia G. Dunbar, Imad M. El-Kebbi, Daniel L. Gallina, Risa P. Hayes, Christopher D. Miller, Mary K. Rhee, Dennis M. Thompson, and Clyde Watkins.** 2005. “An Endocrinologist-Supported Intervention Aimed at Providers Improves Diabetes Management in a Primary Care Site: Improving Primary Care of African Americans with Diabetes (IPCAAD) 7.” *Diabetes Care*, 28(10): 2352–2360.

**Shea, Beverley J., Barnaby C. Reeves, George Wells, Micere Thuku, Candyce Hamel, Julian Moran, David Moher, Peter Tugwell, Vivian Welch, Elizabeth Kristjansson, and David A. Henry.** 2017. “AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both.” *BMJ*, 358.

**Sterne, J. A., and M. Egger.** 2001. “Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis.” *Journal of Clinical Epidemiology*, 54(10): 1046–1055.

**Ward, A, M Kamien, F Mansfield, and B Fatovich.** 1996. "Educational feedback in management of diabetes in general practice." *Education for General Practice*, , (7): 142–150.