

Toward Evidence-Based Quality Improvement

Evidence (and its Limitations) of the Effectiveness of Guideline Dissemination and Implementation Strategies 1966–1998

Jeremy Grimshaw, MBChB, PhD,¹ Martin Eccles, MBChB, MD,² Ruth Thomas, BSc, PhD,³ Graeme MacLennan, BSc, MSc,³ Craig Ramsay, BSc, PhD,³ Cynthia Fraser, MA,³ Luke Vale, MSc, MA³

¹Ottawa Health Research Institute, University of Ottawa, Ottawa, ON, Canada; ²University of Newcastle upon Tyne, Newcastle upon Tyne, UK; ³University of Aberdeen, Aberdeen, UK.

OBJECTIVES: To determine effectiveness and costs of different guideline dissemination and implementation strategies.

DATA SOURCES: MEDLINE (1966 to 1998), HEALTHSTAR (1975 to 1998), Cochrane Controlled Trial Register (4th edn 1998), EMBASE (1980 to 1998), SIGLE (1980 to 1988), and the specialized register of the Cochrane Effective Practice and Organisation of Care group.

REVIEW METHODS: INCLUSION CRITERIA: Randomized-controlled trials, controlled clinical trials, controlled before and after studies, and interrupted time series evaluating guideline dissemination and implementation strategies targeting medically qualified health care professionals that reported objective measures of provider behavior and/or patient outcome. Two reviewers independently abstracted data on the methodologic quality of the studies, characteristics of study setting, participants, targeted behaviors, and interventions. We derived single estimates of dichotomous process variables (e.g., proportion of patients receiving appropriate treatment) for each study comparison and reported the median and range of effect sizes observed by study group and other quality criteria.

RESULTS: We included 309 comparisons derived from 235 studies. The overall quality of the studies was poor. Seventy-three percent of comparisons evaluated multifaceted interventions. Overall, the majority of comparisons (86.6%) observed improvements in care; for example, the median absolute improvement in performance across interventions ranged from 14.1% in 14 cluster-randomized comparisons of reminders, 8.1% in 4 cluster-randomized comparisons of dissemination of educational materials, 7.0% in 5 cluster-randomized comparisons of audit and feedback, and 6.0% in 13 cluster-randomized comparisons of multifaceted interventions involving educational outreach. We found no relationship between the number of components and the effects of multifaceted interventions. Only 29.4% of comparisons reported any economic data.

CONCLUSIONS: Current guideline dissemination and implementation strategies can lead to improvements in care within the context of rigorous evaluative studies. However, there is an imperfect evidence base to support decisions about which guideline dissemination and implementation strategies are likely to be efficient under different circumstances. Decision makers need to use considerable judgment about how best to use the limited resources they have for quality improvement activities.

KEY WORDS: practice guideline; systematic review; implementation research.

DOI: 10.1111/j.1525-1497.2006.00357.x

J GEN INTERN MED 2006; 21:S14–20.

© 2006 by the Authors

Health systems internationally are investing substantial resources in quality improvement (QI) initiatives to promote effective and cost-effective health care. Such initiatives have the potential to improve the care received by patients by promoting interventions of proven benefit and discouraging ineffective interventions. However, individual health care organizations have relatively few resources for QI initiatives and decision makers need to consider how best to use these to maximize benefits. In some circumstances, the costs of QI initiatives are likely to outweigh their potential benefits while in others it may be more efficient to adopt less costly but less effective QI initiatives. In order to make informed judgments, decision makers need to consider a range of factors.¹ First, what are the potential clinical areas for QI initiatives? These may reflect national or local priorities. Decision makers should consider the prevalence of the condition, whether there are effective and efficient health care interventions available to improve patient outcome, and whether there is evidence that current practice is suboptimal. Second, what are the likely benefits and costs required for specific QI strategies? Decision makers need to consider the likely effectiveness of different dissemination and implementation strategies for the targeted condition in their settings as well as the resources required to deliver the different strategies. Third, what are the likely benefits and costs as a result of any changes in provider behavior? In order to answer these questions, decision makers need evidence about the effects of specific QI strategies, the resources needed to deliver them, and how the effects of QI strategies

Address correspondence and requests for reprints to Dr. Grimshaw: Director of the Clinical Epidemiology Program, Ottawa Health Research Institute, Civic Campus, 1053 Carling Avenue, Administration Building, Room 2-017, Ottawa, ON, Canada K1Y 4E9 (e-mail: jgrimshaw@ohri.ca).

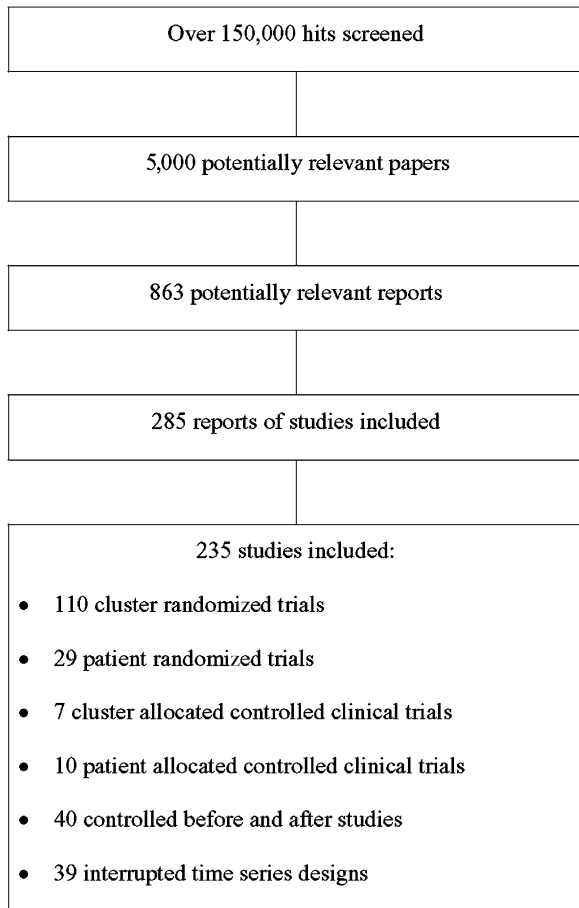


FIGURE 1. Flowchart of included studies.

vary depending on factors such as the context, targeted professionals, and targeted behaviors.

Despite the current interest in QI, uncertainty remains as to the likely effectiveness of different QI strategies and the resources required to deliver them. In 1998, we conducted an overview of 41 published systematic reviews of professional behavior change strategies.² The findings of these reviews suggest that passive dissemination of educational materials (e.g., mailing guidelines to professionals) is largely ineffective, whereas reminders and educational outreach are largely effective and multifaceted interventions are more likely to be effective than single interventions. However, these reviews commonly used vote counting methods that add up the number of positive and negative comparisons to conclude whether or not interventions are effective.³ The vote counting method fails to provide an estimate of the effect size of an intervention (giving equal weight to comparisons that show a 1% change or a 50% change) and ignores the precision of the estimate from the primary comparisons (giving equal weight to comparisons with 100 or 1,000 participants).^{3,4} Further comparisons with potential unit of analysis errors need to be excluded because of the uncertainty of their statistical significance. Underpowered comparisons observing clinically significant but statistically insignificant effects would be counted as “no effect comparisons.”^{3,4} Previous reviews have also tended to describe interventions based on the author’s main description of the intervention, which often ignores co-interventions.

In this paper, we describe the available evidence published between 1966 and 1998 (and highlight its limitations) concerning the effectiveness of clinical practice guideline dissemination and implementation strategies (a common component of QI initiatives) based upon the findings of a systematic review that attempted to address the methodologic weaknesses of previous reviews by adopting a more explicit analytical approach.^{1,5}

SYSTEMATIC REVIEW OF GUIDELINE DISSEMINATION AND IMPLEMENTATION STRATEGIES

The primary aim of the review was to estimate the effectiveness and efficiency of guideline dissemination and implementation strategies to promote improved professional practice. We also determined the frequency and methods of economic appraisal used in included studies and summarized existing evidence on the relative efficiency of guideline dissemination and implementation strategies. The full methods and results of the systematic review are available as a United Kingdom NHS Health Technology Assessment monograph.^{1,5} We briefly summarize the methods below.

Methods

We searched MEDLINE, EMBASE, Health Star, the Cochrane Controlled Trials Register, and System for Information on Grey Literature in Europe (SIGLE) using a highly sensitive search strategy developed for the Cochrane-Effective Practice and Organisation of Care (EPOC) group between 1976 and 1998.⁵ Searches were not restricted by language or publication type. We included cluster and individual randomized-controlled trials (RCTs), controlled clinical trials, controlled before and after studies (CBAs), and interrupted time series (ITS) that evaluated any guideline dissemination or implementation strategy targeting physicians and that reported an objective measure of provider behavior and/or patient outcome. Two reviewers independently screened the search results, assessed studies against the inclusion criteria, and abstracted data from the included studies. Disagreements were resolved by consensus. Data were abstracted concerning study design, methodologic quality (using the EPOC quality appraisal criteria⁵), characteristics of participants, study settings, and targeted behaviors, characteristics of the interventions (using the EPOC taxonomy, see Table 1) and study results. Studies reporting any economic data were also assessed against the *British Medical Journal* guidelines for reviewers of economic evaluations.⁶

We used absolute improvement in dichotomous process of care measures (e.g., proportion of patients receiving appropriate treatment) as the primary effect size for each comparison for 2 pragmatic reasons. First, dichotomous process of care measures were reported considerably more frequently in the studies. Second, there were problems with the reporting of continuous process of care variables. We initially planned to calculate standardized mean differences for continuous process of care measures. However, few studies reported sufficient data to allow this. We considered calculating the relative percentage change; however, we considered this approach relatively uninformative. For example, a relative percentage change in a continuous measure depends on the scale being used—a comparison that shifts from a mean of 1 to 2 will show

Table 1. Classification of Professional Interventions from EPOC Taxonomy

-
-
- (a) *Distribution of educational materials*—distribution of published or printed recommendations for clinical care, including clinical practice guidelines, audio-visual materials, and electronic publications
 - (b) *Educational meetings*—health care providers who have participated in conferences, lectures, workshops, or traineeships
 - (c) *Local consensus processes*—inclusion of participating providers in discussion to ensure that they agreed that the chosen clinical problem was important and the approach to managing the problem was appropriate
 - (d) *Educational outreach visits*—use of a trained person who met with providers in their practice settings to give information with the intent of changing the provider's practice
 - (e) *Local opinion leaders*—use of providers nominated by their colleagues as “educationally influential.” The investigators must have explicitly stated that their colleagues identified the opinion leaders
 - (f) *Patient mediated interventions*—new clinical information (not previously available) collected directly from patients and given to the provider, e.g., depression scores from an instrument
 - (g) *Audit and feedback*—any summary of clinical performance of health care over a specified period of time
 - (h) *Reminders*—patient or encounter-specific information, provided verbally, on paper or on a computer screen that is designed or intended to prompt a health professional to recall information
 - (i) *Marketing*—use of personal interviewing, group discussion (“focus groups”), or a survey of targeted providers to identify barriers to change and subsequent design of an intervention that addresses identified barriers
 - (j) *Mass media*—(i) varied use of communication that reached great numbers of people including television, radio, newspapers, posters, leaflets, and booklets, alone or in conjunction with other interventions; and (ii) targeted at the population level
-

EPOC. Cochrane-Effective Practice and Organisation of Care.

the same relative improvement as one that shifts from 25 to 50. Where studies reported more than 1 measure of each end point, we abstracted the primary measure (as defined by the authors of the study) or the median measure. For example, if the comparison reported 5 dichotomous process of care variables and none of them were denoted as the primary variable then we would rank the effect sizes for the 5 variables and take the median value. We standardized effect sizes so that a positive difference between postintervention percentages or means was a good outcome. We tried to reanalyze cluster-allocated studies with unit of analysis errors (e.g., studies that randomized health care practices of physicians but analyzed patient-level data⁵) if the authors provided cluster-level data or an estimate of the observed intraclass correlation. We reanalyzed ITS that had been inappropriately analyzed (e.g., ITS that were analyzed as simple before and after studies) using time series regressions.^{5,7}

Given the expected extreme heterogeneity within the review and the number of studies with potential unit of analysis errors, we did not plan to undertake formal meta-analysis. For all designs (other than ITS), we reported (separately for each study design): the number of comparisons showing a positive direction of effect; the median effect size across all comparisons; the median effect size across comparisons without unit of analysis errors; and the number of comparisons showing statistically significant effects. This allowed the reader to assess the consistency of effects across different study designs and across comparisons where the statistical significance is known.

Description of Included Studies

The search strategy produced over 1,50,000 hits. The titles and abstracts of these were screened and around 5,000 were initially identified as potentially relevant. The full text of 863 potentially relevant reports of studies were retrieved and assessed for inclusion in this review. We excluded 628 studies because the intervention or design did not meet our inclusion criteria. The review included 235 studies (Fig. 1), including 110 (46.8%) cluster-randomized trials (C-RCTs), 29 (12.3%) patient-randomized trials (P-RCTs), 7 (3.0%) cluster allocated controlled clinical trials (C-CCTs), 10 (4.3%) patient allocated controlled clinical trials (P-CCT), 40 (17.0%) CBAs, and 39

(16.6%) ITS designs. Overall, the quality of studies was difficult to determine because of poor reporting. Unit of analysis errors were observed in 53% of C-RCTs, 86% of C-CCTs, and 83% of CBAs. It was only possible to reanalyze 1 C-RCT using data provided within the original paper. The studies were relatively small; studies that allocated to study arms by cluster (C-RCTs, C-CCTs, CBAs) had a median number of 7 units per arm (interquartile range 1 to 24). The median number of data points before and after the intervention for ITS studies were 10 (interquartile range 5 to 17) and 12 (interquartile range 7 to 24), respectively. Only 41% of the ITS studies appeared to have been analyzed correctly.

The studies were conducted in 14 different countries, most commonly the United States (71%). The most common setting was primary care (39%), followed by “inpatient” settings (19%) and generalist outpatient or “ambulatory care” settings (19%).

Process of care was measured in 95% of studies; 67% of studies reported dichotomous process measures (primary end point). Seventy-nine percent of studies involved only 1 comparison of an intervention group versus control group, 13% involved 2 comparisons (e.g., 3-arm RCT), and 8% involved 3 or more comparisons. As a result, the review included 309 comparisons. Eighty-four of the 309 (27%) comparisons involved a study group receiving a single guideline implementation intervention strategy versus a “no intervention” or “usual care” control group. The most frequent single intervention evaluated against a “no intervention” control was reminders in 13% of all comparisons, followed by dissemination of educational materials in 6% of comparisons, audit and feedback in 4% of comparisons, and patient-directed interventions in 3% of comparisons. One hundred and seventeen studies (including 136 comparisons) evaluated a total of 68 different combinations of interventions against a “no intervention” control group and 61 studies (including 85 comparisons) evaluated 58 different combinations of interventions against a control that also received 1 or more intervention. The intervention strategy used most frequently as part of multifaceted interventions was educational materials (evaluated in 48% of all comparisons), followed by educational meetings (41% of all comparisons), reminders (31% of all comparisons), and audit and feedback (24% of all comparisons).

Results

For the purposes of this paper, we highlight a number of key findings. Overall, the majority of comparisons reporting dichotomous process data (86.6%) observed improvements in care, suggesting that dissemination and implementation of guidelines can promote compliance with recommended practices.

Single Interventions

Educational Materials. Five comparisons reported dichotomous process data including 4 C-RCTs and 1 P-RCT comparison (A76). All C-RCT comparisons observed improvements in care; the median effect was +8.1% (range +3.6% to +17%) absolute improvement in performance. Two comparisons had potential unit of analysis errors and the significance of 1 comparison could not be determined. The remaining comparison without a potential unit of analysis error observed an effect of +6% (not significant). The P-RCT comparison observed an -8.3% absolute deterioration in care (not significant).

Audit and Feedback. Six comparisons reported dichotomous process data including 5 C-RCTs and 1 CBA. All 5 C-RCT comparisons observed improvements in care. Across all comparisons, the median effect was +7.0% (range +1.3% to +16.0%) absolute improvement in performance. Three comparisons had potential unit of analysis errors. The 2 remaining comparisons observed effects of +5.2% (not significant) and +13% ($P < .05$). The CBA comparison observed an absolute improvement in performance of +32.6%; this study had a potential unit of analysis error.

Reminders. Thirty-three comparisons reported dichotomous process data including 15 C-RCTs, 8 P-RCTs, 8 CCTs, and 2 CBAs. Twelve of 14 C-RCT comparisons reported improvements in care; the median effect was +14.1% (range -1.0% to +34.0%) absolute improvement in performance. Eleven comparisons had potential unit of analysis errors; the remaining 3 comparisons observed a median effect of +20.0% (range +13% to +20%), all were significant. Comparable data could not be abstracted for 1 C-RCT comparison, which reported no significant changes in overall compliance and had a potential unit of analysis error. Seven of 8 P-RCT comparisons reported improvements in care; the median effect was +5.4% (range -1% to +25.7%). Three comparisons were statistically significant and 1 study had a potential unit of analysis error. One C-CCT reported an absolute improvement in care of +4.3% but had a potential unit of analysis error. Six of the 7 P-CCT comparisons reported improvements in care; the median effect was +10.0% (range 0% to +40.0%) absolute improvement in care. Four comparisons were statistically significant. Two CBA comparisons observed effects of +3.6% and +10% absolute improvements in performance; both had potential unit of analysis errors.

Multifaceted Interventions

We originally planned to undertake a metaregression analysis to estimate the effects of different component interventions; however, the extreme number of combinations of multifaceted studies proved problematic. Within the full report, we summarize the results of all the multifaceted interventions. Here, we highlight 2 sets of results—the effects of multifaceted interventions including educational outreach

and an analysis of whether the effectiveness of multifaceted interventions increases with the number of component interventions.

Multifaceted Interventions Including Educational Outreach. Eighteen comparisons reported dichotomous process data including 13 C-RCTs and 5 CBAs. Eleven of the C-RCT comparisons observed improvements in performance; the median effect was +6.0% (range -4% to +17.4%) absolute improvement in performance. Statistical significance could be determined for 5 comparisons; the median effect size across these studies was +10.0% (range -4% to +17.4%) absolute improvement in performance (only 1 study observing a +17.4% absolute improvement in performance was statistically significant). Seven studies had potential unit of analysis errors and the significance of the postintervention comparison could not be determined in 1 study. Two of 4 CBA comparisons reporting dichotomous process of care results observed positive improvements in performance. Across all studies the median effect was +7.3 (range -5.6% to +16.4%) absolute improvement in performance. All these studies had potential unit of analysis errors. Comparable data could not be abstracted for 1 study that reported significant improvements in use of antibiotic prophylaxis in surgery.

Does the Effectiveness of Multifaceted Interventions Increase with the Number of Interventions? Previous systematic reviews have suggested that the effectiveness of multifaceted interventions increases with the number of component interventions; in other words, there is a dose-response curve. Figure 2 illustrates the spread of effect sizes for increasing number of interventions in the study group using box plots. Visually, there appeared to be no relationship between effect size and number of interventions. There was no statistical evidence of a relationship between the number of interventions used in the study group and the effect size (Kruskal-Wallis test, $P = .18$).

Economic Evaluations

Only 29% of studies reported any economic data. Eleven reported cost-effectiveness analyses, 38 reported cost-consequence analyses (where differences in cost were set against differences in several measures of effectiveness), and 14 re-

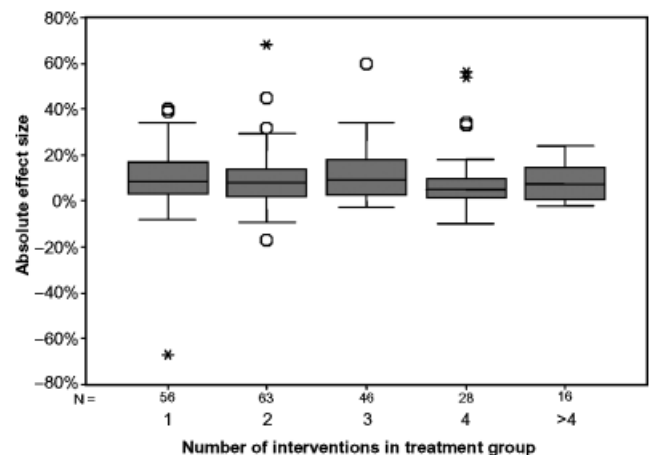


FIGURE 2. Effect sizes of multifaceted interventions by number of component interventions.

ported cost analyses (where some aspect of cost was reported but not related to benefits). The majority of studies only reported costs of treatment; only 25 studies reported data on the costs of guideline development or guideline dissemination and implementation. The majority of studies used process measures for their primary end point despite the fact that only 3 guidelines were explicitly evidence based (and may not have been efficient).

Overall, the methods of the economic evaluations and cost analyses were poor. The viewpoint adopted in economic evaluations was only stated in 10 studies. The methods to estimate costs were comprehensive in about half of the studies, and few studies reported details of resource use. Because of the poor quality of reporting of the economic evaluation, data on resource use and cost of guideline development, dissemination, and implementation were not available for most of them. Only 4 studies provided sufficiently robust data for consideration. These studies demonstrated that the costs of local guideline development are not insubstantial if explicitly costed, although they recognized that the time needed for many activities is frequently not made explicit and that these activities are often undertaken outside work time. Further estimates of the resources and costs of different dissemination and implementation are needed before the generalizability of the reported costs can be determined.

DISCUSSION

Strengths and Weaknesses

The systematic review is the most extensive, methodologically sophisticated, and up-to-date review of clinical guideline implementation strategies. We developed highly sensitive search strategies and consider the included studies to be a relatively complete set of studies for the period 1966 to mid-1998. We undertook detailed data abstraction about the quality of the studies, characteristics of the studies, and interventions. As a result, the methodologic weaknesses of the primary studies have been made more explicit within this review. Where possible, we attempted to reanalyze the results of comparisons with common methodologic errors. We have also characterized the interventions in greater detail than previous reviews; this revealed the multifaceted nature of the majority of evaluated interventions. We used a more explicit analytical framework than previous reviews that allowed us to consider the methodologic quality of studies (based on design and presence of unit of analysis errors when interpreting the results of comparisons). Our approach focuses on the observed effect sizes and does not consider statistical significance (because of problems of estimating statistical significance in studies with unit of analysis errors) or weight by study size (because of problems of estimating effective sample size in studies with unit of analysis errors). However, it did allow us to provide some information about the potential effect size of interventions, which we believe is more informative than previous reviews using vote-counting approaches. More recently, reviews have weighted effects by the total number of professionals involved⁸ or of effective patient sample size using externally imputed intra-class correlations for studies with unit of analysis errors.⁹ The optimal method of weighting of cluster-allocated studies (especially those with unit of analysis errors) remains unclear. It is also likely that our review suffers from publication bias

although the extent of publication bias has been poorly studied in rigorous evaluations of QI strategies (however, a recent review has suggested that smaller nonrandomized studies of diabetes QI interventions led to larger improvements in diabetic control⁹). As a result, the reader should consider the likelihood that this review's findings are over optimistic. Finally, our review only considered studies published until 1998. We are aware that this is a very active research field and would expect between 30 and 50 new papers to be published each year. Further, new QI interventions are being developed constantly (although rigorous evaluation often lags considerably behind). This demonstrates the significant logistical challenge of undertaking a systematic review in this area. Nevertheless, our unsystematic impression is that the field has not become dramatically clearer since our review was undertaken.

Key Findings

Key findings of the review are as follows. The majority of comparisons reporting dichotomous process data (86.6%) observed improvements in care, suggesting that dissemination and implementation of guidelines can promote compliance with recommended practices. Second, reminders are a potentially effective intervention and are likely to result in moderate improvements in process of care. The majority of comparisons evaluated either paper-based reminders or computer-generated paper-based reminders across a wide range of settings and targeted behaviors. Third, educational outreach was often a component of a multifaceted intervention and might be considered to be inherently multifaceted. The results suggest that the educational outreach may result in modest improvements in process of care but this needs to be offset against both the resources required to achieve this change and practical considerations. Fourth, educational materials and audit and feedback appeared to result in modest effects. Finally, although the majority of comparisons evaluated multifaceted interventions, there were few replications of either specific multifaceted interventions against a no intervention control or against a specific control group. Across all combinations, multifaceted interventions did not appear to be more effective than single interventions and the effects of multifaceted interventions did not appear to increase with the number of component interventions.

Comparison with Previous Systematic Review Findings

This review has arrived at conclusions that may appear at odds with the overview and 2 previous systematic reviews. For example, Freemantle et al.¹⁰ reviewed 11 studies evaluating the effects of the dissemination of educational materials. They used vote counting methods and excluded studies with unit of analysis errors. None of the comparisons using appropriate statistical techniques found statistically significant improvements in practice. In this review, we used a more explicit analytical framework to explore the median and range of effect sizes in studies with and without unit of analysis errors. We identified 4 C-RCT comparisons reporting dichotomous process measures. All of these observed improvements in care; the median effect was +8.1% (range +3.6% to +17%) absolute improvement in performance. Two comparisons had potential unit of analysis errors and the significance of 1 comparison could not be determined. The remaining comparison without a

potential unit of analysis error observed an effect of +6% (not significant). If we had used the same analytical approach, we would have reached similar conclusions to Freemantle et al.¹⁰ However, such an approach would have failed to detect that printed educational materials led to improvements in care across the majority of studies (albeit that the statistical significance of the majority of comparisons is uncertain). Based upon this review, we would not conclude that printed educational materials are effective given the methodologic weaknesses of the primary studies. Instead, we conclude that printed educational materials may lead to improvements in care and that policy makers should not dismiss printed educational materials given their possible effect, relative low cost, and feasibility within health service settings.

Previous reviews have also suggested that multifaceted interventions are more effective than single interventions on the basis that they address multiple barriers to implementation. Davis et al.'s¹¹ review of continuing medical education strategies concluded that multifaceted interventions were more likely to be effective. Wensing et al.¹² undertook a review of the effectiveness of introducing guidelines in primary care settings; they concluded that multifaceted interventions combining more than 1 intervention tended to be more effective but might be more expensive. The specific details about how interventions were coded and the analytical method of these 2 reviews are unclear. In this review, we coded all intervention components and used explicit methods to determine a single effect size for each study. The analysis suggested that the effectiveness of multifaceted interventions did not increase incrementally with the number of components. Few studies provided any explicit rationale or theoretical base for the choice of intervention. As a result, it was unclear whether researchers had an a priori rationale for the choice of components in multifaceted interventions based upon possible causal mechanisms or whether the choice was based on a "kitchen sink" approach. It is plausible that multifaceted interventions built upon a careful assessment of barriers and coherent theoretical base may be more effective than single interventions.

Implications for Decision Makers

Decision makers need to use considerable judgment about which interventions are most likely to be effective in any given circumstance and choose intervention(s) based upon consideration of the feasibility, costs, and benefits potentially effective interventions are likely to yield. Wherever possible, interventions should include paper-based or computerized reminders. It may be more efficient to use a cheaper more feasible but less effective intervention (for example, passive dissemination of guidelines) than a more expensive but potentially more effective intervention. Decision makers also need to consider the resource implications associated with consequent changes in clinical practice and assess their likely impact on different budgets.

Decision makers should assess the effects of any interventions preferably using rigorous evaluative designs.

Implications for Further Research

This review highlights the fact that despite 30 years of research in this area, we still lack a robust, generalizable evidence base to inform decisions about QI strategies.

The lack of a coherent theoretical basis for understanding professional and organizational behavior change limits our ability to formulate hypotheses about which interventions are likely to be effective under different circumstances and hampers our understanding of the likely generalizability of the results of definitive trials. An important focus for future research should be to develop a better theoretical understanding of professional and organizational behavior change to allow us to explore mediators and moderators of behavior change.¹³

As in other areas of medical and social research, rigorous evaluations (mainly RCTs) will provide the most robust evidence about the effectiveness of guideline dissemination and implementation strategies because effects are likely to be modest, there is substantial risk of bias, and we have poor understanding of the likely effect modifiers and confounders. However, despite the considerable number of evaluations that have been undertaken in this area, we still lack a coherent evidence base to support decisions about which dissemination and implementation strategies to use. This is partly because of the poor methodologic quality of the existing studies; for example, the statistical significance of many comparisons could not be determined from the published studies because of common methodologic weaknesses. Further rigorous evaluations of different dissemination and implementation strategies need to take place. These evaluations need to be methodologically robust (addressing the common methodologic errors identified in the systematic review), incorporate economic evaluations and, wherever possible, explicitly test behavioral theories to further develop our theoretical understanding of factors influencing the effectiveness of guideline dissemination and implementation strategies.

Other members of the Guideline dissemination and implementation systematic review team include Cam Donaldson and Paula Whitty (Centre for Health Services Research, University of Newcastle upon Tyne, UK), Lloyd Matowe, Liz Shirran (Health Services Research Unit, University of Aberdeen, UK), and Michel Wensing and Rob Dijkstra (Centre for Quality of Care Research, University of Nijmegen, the Netherlands). The project was funded by the UK NHS Health Technology Assessment Program. The Health Services Research Unit and Health Economics Research Unit are funded by the Chief Scientist Office of the Scottish Executive Department of Health. Jeremy Grimshaw holds a Canada Research Chair in Health Knowledge Transfer and Uptake. Ruth Thomas is a Wellcome Training Fellow in Health Services Research. The views expressed are those of the authors and not necessarily the funding bodies.

The project was funded by the UK NHS Health Technology Assessment Program. The sponsor had no involvement in study design, in the collection, analysis, and interpretation of data.

On behalf of the UK NHS HTA GUIDELINE dissemination and implementation systematic review team.

REFERENCES

1. **Grimshaw JM, Thomas RE, MacLennan G, et al.** Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technol Assessment*. 2004;8:1-72. Available at: <http://www.hta.nhsweb.nhs.uk>. Accessed August 5, 2004.
2. **Grimshaw JM, Shirran L, Thomas RE, et al.** Changing provider behaviour: an overview of systematic reviews of interventions. *Med Care*. 2001;39(suppl 2):II-2-II-45.
3. **Bushman BJ.** Vote counting methods in meta-analysis. In: Cooper H, Hedges L, eds. *The Handbook of Research Synthesis*. New York: Russell Sage Foundation; 1994.

4. **Grimshaw J, McAuley L, Bero L, et al.** Systematic reviews of the effectiveness of quality improvement strategies and programs. *Qual Safety Health Care.* 2003;12:293–303.
5. **Grimshaw JM, Thomas RE, MacLennan G, et al.** Effectiveness and efficiency of guideline dissemination and implementation strategies. *Appendices. Health Technol Assessment.* 2004;8:73–309. Available at: <http://www.hta.nhsweb.nhs.uk>. Accessed August 5, 2004.
6. **Drummond MF, Jefferson TO,** for the BMJ working party on guidelines for authors and peer-reviewers of economic submissions to the *British Medical Journal.* Guidelines for authors and peer-reviewers of economic submissions to the *British Medical Journal.* *BMJ.* 1996;313:275–83.
7. **Ramsay C, Matowe L, Grilli R, Grimshaw J, Thomas R.** Interrupted time series designs in health technology assessment: lessons from two systematic reviews of behaviour change strategies. *Int J Technol Assessment Health Care.* 2003;19:613–23.
8. **Jamtvedt G, Young JM, Kristoffersen DT, Thomson O'Brien MA, Oxman AD.** Audit and feedback: effects on professional practice and health care outcomes. *The Cochrane Database of Systematic Reviews,* 2003, Art. No.: CD000259. DOI: 10.1002/14651858.CD000259.
9. **Shojania KG, Ranji SR, Shaw LK, et al.** *Diabetes Mellitus Care.* Vol. 2 of Shojania KG, McDonald KM, Wachter RM, Owens DK. *Closing The Quality Gap: A Critical Analysis of Quality Improvement Strategies.* Technical Review 9 (Contract No. 290-02-0017 to the Stanford University–UCSF Evidence-Based Practice Center). AHRQ Publication No. 04-0051-2. Rockville, MD: Agency for Healthcare Research and Quality; September 2004.
10. **Freemantle N, Harvey EL, Wolf F, Grimshaw JM, Grilli R, Bero LA.** Printed educational materials to improve the behaviour of health care professionals and patient outcome (Cochrane Review). In: *The Cochrane Library,* Issue 4, 1996 Oxford: Update software
11. **Davis DA, Thomson MA, Oxman AD, Haynes RB.** Changing physician performance. A systematic review of the effect of continuing medical education strategies. *JAMA.* 1995;274:700–5.
12. **Wensing M, Van der Weijden T, Grol R.** Implementing guidelines and innovations in general practice: which interventions are effective? *J Gen Practice.* 1998;48:991–7.
13. **Eccles M, Grimshaw J, Walker A, Johnston M, Pitts N.** Changing the behaviour of healthcare professionals: the use of theory in promoting the uptake of research findings. *J Clin Epidemiol.* 2005;58:107–12.