Chapter 2. Evidence-based Review Methodology for the *Closing the Quality Gap* Series

Kaveh G. Shojania, M.D.
University of California, San Francisco

Kathryn M. McDonald, M.M.
Stanford University

Douglas K. Owens, M.D., M.S.
VA Palo Alto Health Care System
and Stanford University

Definition and Scope

The Stanford University–UCSF Evidence-based Practice Center (EPC) performed a comprehensive review of the evidence relating to a broad range of quality improvement (QI) strategies and their utility in a variety of clinical areas. The topic areas were chosen from a group of 20 priority conditions identified by the IOM[1] (see Appendix A). For this project, the authors defined the following terms:

Quality of health care: The degree to which health services for individuals and patient populations increase the likelihood of desirable health outcomes and are consistent with current professional knowledge.[47]

Quality gap: The difference between health care processes or outcomes observed in practice, and those thought to be achievable with the most current and effective professional knowledge. The difference must be attributable in whole or in part to a deficiency that could be addressed by the health care system. An example of a process-level quality gap for hypertension involves the 62 percent of clinical visits during which physicians failed to introduce evidence-based, guideline-concordant drug therapy to patients with a systolic blood pressure of 140mm/HG or higher.[48]

An example of an outcome-level quality gap for myocardial infarctions involves a disparity in survival rates. Despite numerous new therapies that have substantially decreased mortality over the past 25 years, survival gains have occurred mainly in males and in younger patients, with less gain in women and the elderly.[49] The resolution of such outcome gaps generally entails detailed analyses of relevant treatment and care processes, in an effort to explain their genesis and identify targets for action.

Quality improvement (QI) strategy: Any intervention aimed at reducing the quality gap for a group of patients representative of those encountered in routine practice. For the purposes of their literature search, the authors considered a study to include a QI strategy evaluation if any of the following applied:

- The intervention targeted implementation of a particular process of care (or set of processes) believed to benefit patients with the priority condition(s); i.e., interventions

designed to improve provider adherence to a clinical best practice guideline, or those intended to increase the proportion of patients who received recommended care.

- The intervention targeted implementation of a structural or organizational feature believed to benefit patients with the priority condition; i.e., interventions that changed the care provider, added supplemental personnel, or made clinical information systems part of the treatment protocol.
- The intervention attempted to improve outcomes for a broad and relatively unselected group of patients with the priority condition; i.e., interventions designed to improve the delivery of care for all patients with diabetes or hypertension at a specific clinic.
- The intervention targeted a subset of patients that typically is excluded from clinical research; i.e., frail elders, minorities, the economically disadvantaged, or those with multiple comorbid conditions.
- The intervention involved any of the specific QI strategies falling within a taxonomy of approaches to QI that the authors developed, based on evaluations of various quality improvement interventions[50-59] and authoritative definitions[60] (see below for taxonomy).

Quality improvement target: The outcome, process, or structure that the QI strategy is intended to influence, with the goal of reducing the quality gap. A target typically would be a measure of disease control, including direct health outcomes (morbidity or mortality), or intermediate outcomes proven to influence direct outcomes (such as blood pressure or hemoglobin A_{1C} control). Targets also may involve adherence to accepted processes of care, either by clinicians (i.e., guideline recommendations and performance measures) or by patients (i.e., adherence to prescribed medications, recommended self-management).

Taxonomy of Quality Improvement Strategies

To ensure consistency in their review and evaluation of the literature, the authors developed a taxonomy that modifies several well-established classification systems.[50-54] A recent and systematic review of disease management studies combined QI strategies and targets, classifying interventions as: provider education, provider feedback, provider reminders, patient education, patient reminders, and patient financial incentives.[54] The Cochrane EPOC data collection instrument uses four broad classifications (professional interventions, organizational interventions, financial interventions, and regulatory interventions), each of which has detailed subcategories. An alternative taxonomy, described in a recent systematic review of interventions to promote immunization and cancer screening,[52] specifies three dimensions for characterizing QI strategies – the *type* of QI strategy (e.g., education, audit and feedback, organizational changes, financial incentives), *mediators* of the intervention (e.g., use of local opinion leaders, involvement of top management, identification of barriers to change), and *target audience* (e.g., patients, providers, health care delivery systems). The authors of this series modified the various taxonomies to better facilitate their review of the evidence.

Types of QI Strategies

Nine types of QI strategies are outlined below, along with key substrategies. These categories are broad, and, in some cases, combine multiple interventions. The authors explored this heterogeneity in their analyses to assess the possibility of making inferences and judgments about the success of the strategy as a whole, or whether further subdivision would be needed.

Where relevant, the analyses also take into consideration the fact that many interventions are multifaceted and employ more than one type of QI strategy.

1. **Provider reminder systems**—the investigators defined a reminder system as any patient- or clinical encounter-specific information, provided verbally, in writing, or by computer, to prompt a clinician to recall information, or intended to prompt consideration of a specific process of care (i.e., "This patient last underwent screening mammography 3 years ago"). The reminder also may include information prompting the clinician to follow evidence-based care recommendations (e.g., to make medication adjustments, or to order appropriate screening tests). The phrase "clinical encounter-specific" in the definition serves to distinguish reminder systems from audit and feedback, where clinicians typically receive performance summaries relative to a process or outcome of care spanning multiple encounters (e.g., all Type 2 diabetic patients seen by the clinician during the past 6 months).

2. **Facilitated relay of clinical data to providers**—used to describe the transfer of clinical information collected directly from patients and relayed to the provider, in instances where the data are not generally collected during a patient visit, or using some format other than the existing local medical record system (i.e., the telephone transmission of a patient's blood pressure measurements, from a specialist's office). The EPOC group uses the term "patient mediated" to describe such interventions, but the authors regard the above label as more descriptive. Some overlap with provider reminder systems was expected, but the strategies were kept separate at the abstraction stage. This decision allowed for the possibility that the data could be subsequently analyzed with and without collapsing the two strategies.

3. **Audit and feedback**—the researchers defined audit and feedback as any summary of clinical performance for health care providers or institutions, performed for a specific period of time and reported either publicly or confidentially to the clinician or institution (e.g., the percentage of a provider's patients who achieved or did not achieve some clinical target, such as blood pressure or HbA_{1c} control over a certain period). Benchmarking is a term referring to the provision of performance data from institutions or providers regarded as leaders in the field. These data serve as performance targets for other providers and institutions. The authors included benchmarking as a type of audit and feedback, so long as local data were provided for comparison with the benchmark data.

4. **Provider education**—used to describe a variety of interventions including educational workshops, meetings (e.g., traditional Continuing Medical Education [CME]), lectures (in person or computer-based), educational outreach visits (by a trained representative who meets with providers in their practice settings to disseminate information with the intent of changing the providers' practice). The same term also is used to describe the distribution of educational materials (electronically published or printed clinical practice guidelines and audio-visual materials). The investigators further captured information about the intensity (i.e., duration and number of educational sessions) and format (i.e., lectures delivered live, via teleconference, or pre-recorded) in a free-text mode, for each of these substrategies. Early plans to capture these and other predictors in a structured form were abandoned after the authors and their technical advisors agreed the judgments

were too subjective. This was due in large part to a relative lack of detail surrounding the interventions in the vast majority of studies.

5. **Patient education**—this strategy is centered on in-person patient education, either individually or as part of a group or community, and through the introduction of print or audio-visual educational materials. Patient education may be the sole component of a particular quality improvement strategy, or it can be one part of a multifaceted QI strategy. It should be noted that the authors evaluated only those strategies in which patient education was regarded as one component of a multifaceted strategy. A future volume in this series may address the topic of patient education as a singular intervention, along with its relative effects on a variety of chronic diseases.

6. **Promotion of self-management**—this strategy includes the distribution of materials (i.e., devices for blood pressure or glucose self-monitoring) or access to a resource that enhances the patients' ability to manage their condition, the communication of useful clinical data to the patient (e.g., most recent HbA_{1c} or lipid panel levels), or followup phone calls from the provider to the patient, with recommended adjustments to care. The authors expected some overlap with regard to patient education (strategy 5) and patient reminders (strategy 7). They elected to keep the strategies separate at the abstraction stage, to allow for the possibility that the data could be analyzed after the fact, with and without collapsing the two strategies.

7. **Patient reminders**—used to define any effort directed by providers toward patients that encourages them to keep appointments or adhere to other aspects of the self-management of their condition.

8. **Organizational change**—this strategy included any intervention having features consistent with at least one of the following descriptions, each of which represents a substrategy of organizational change that was abstracted for incorporation in the analysis:

   a) Disease management or case management – the coordination of assessment, treatment, and referrals by a person or multidisciplinary team in collaboration with, or supplementary to, the primary care provider.

   b) Team or personnel changes – adding new members to a treatment team (e.g., the addition of a diabetes nurse, a clinical pharmacist, or a nutritionist to a clinical practice), creating multidisciplinary teams within a practice, or revising the roles of existing team members (e.g., a clinic nurse is given a more active role in patient management), or the simple addition of more nurses, pharmacists, or physicians to a clinical setting.

   c) Communications, case discussions, and the exchange of treatment information between distant health professionals (i.e., telemedicine).

   d) Total Quality Management (TQM) or Continuous Quality Improvement (CQI) techniques for measuring quality problems, designing interventions and their implementation, along with process re-measurements.

   e) Changes in medical records systems—adopting improved office technology (e.g., computer-based records, patient tracking systems).

Although the definition used for this strategy is consistent with prior reviews,[52] the authors recognized the potential heterogeneity of included interventions and accordingly planned to analyze this strategy with respect to the aforementioned substrategies.

9. **Financial, regulatory or legislative incentives—**this strategy encompassed any intervention having features consistent with at least one of the following descriptions:
   a) Positive or negative financial incentives directed at providers (e.g., regarding adherence to some process of care or achievement of target patient outcome).
   b) Positive or negative financial incentives directed at patients.
   c) System-wide changes in reimbursement (e.g., capitation, prospective payment, shift from fee-for-service to salary).
   d) Changes to provider licensure requirements.
   e) Changes to institutional accreditation requirements.

The authors further abstracted information about the use of clinical information systems, including their role in identifying eligible study participants for QI interventions, for generating clinical reminders, for enabling decision support, and their ability to cultivate data for audit and feedback.

Table 1 presents the major types of QI strategies in the first column, with examples of corresponding substrategies in the second column. The table illustrates the manner in which some QI strategies and substrategies target a single audience, while others attempt to influence multiple audiences, such as patients and health care delivery systems. Many QI strategies evaluated in the literature combine substrategies and audience targets, a situation that makes for challenging analyses of effectiveness. Such combinations often limit the ability of researchers to interpret the active component(s) of a particular intervention.

Table 1. Taxonomy of QI strategies with examples of substrategies

QI strategy	Examples
Provider reminder systems	• Reminders in charts for providers • Computer-based reminders for providers • Computer-based decision support
Facilitated relay of clinical data to providers	• Transmission of clinical data from outpatient specialty clinic to primary care provider by means other than medical record, e.g., phone call or fax
Audit and feedback	• Feedback of performance to individual providers • Quality indicators and reports • National/State quality report cards • Publicly released performance data • Benchmarking – provision of outcomes data from top performers for comparison with provider's own data

QI strategy	Examples
Provider education	• Workshops and conferences • Educational outreach visits (e.g., academic detailing) • Distributed educational materials
Patient education	• Classes • Parent and family education • Patient pamphlets • Intensive education strategies promoting self-management of chronic conditions
Promotion of self-management	• Materials and devices promoting self-management
Patient reminder systems	• Postcards or calls to patients
Organizational change	• Case Management, Disease Management • TQM, CQI techniques • Multidisciplinary teams • Change from paper to computer-based records • Increased staffing • Skill mix changes
Financial incentives, regulation, and policy	*Provider-Directed:* • Financial incentives based on achievement of performance goals • Alternative reimbursement systems (e.g., fee-for-service, capitated payments) • Licensure requirements *Patient-Directed:* • Co-payments for certain visit types • Health insurance premiums, user fees *Health System-Directed:* • Initiatives by accreditation bodies (e.g., residency work hour limits) • Changes in reimbursement schemes (e.g., capitation, prospective payment, salaried providers)

Identification of Quality Improvement Strategies for Evaluation

The medical conditions selected for evaluation were taken from the IOM National Priorities report,[1] and were based on the priorities of stakeholders, the quality of evidence in relation to the usefulness of QI strategies, the expertise of the EPC, and available resources. As described in the Introduction to this volume, the selected topics will be analyzed in a series of volumes to be published over the course of the next two years. The final volume may be used to examine crosscutting analyses of selected QI strategies for many of the disease topics presented in the series.

Search Strategy

The authors initially reviewed QI strategies for hypertension and diabetes to help formulate their methodologic approach. They searched the MEDLINE® database from 1980-present, the Cochrane databases, and the Cochrane registry for the selected topics. The general search strategy was consistent across these two topics. Appendix B illustrates the search strings for hypertension. They searched terms relevant to care coordination and disease management, quality improvement (including Total Quality Management and Continuous Quality Improvement), continuing medical education, educational outreach, audit and feedback, financial incentives, information technologies, telemedicine, and the specific condition under consideration (e.g., hypertension). Additional searches were undertaken for systematic reviews and manual searches also were done, when appropriate, for relevant references. The bibliographies of all articles that met final inclusion criteria were scanned by hand for the project, as were the bibliographies for all relevant systematic reviews and meta-analyses. In cases where no systematic review was found to exist for a given topic, the authors searched the bibliographies of traditional (narrative) review articles, editorials, and news items that appeared to describe QI studies involving outpatient diabetic care.

These searches were supplemented with reviews of citations from the Cochrane EPOC registry of quality improvement strategies. Each of the Collaborative Review Groups within the Cochrane Collaboration works to prepare and maintain systematic reviews of the prevention, treatment, and rehabilitation of a particular health problem or groups of problems, known as the 'scope' of the group. EPOC's mandate is the systematic review of educational, behavioral, financial, organizational, and regulatory interventions designed to improve health professional practice and the organization of health care services, using the most statistically reliable methods, and across all clinical areas.

The EPOC registry has been developed using a highly sensitive search strategy to identify studies within EPOC's scope. The registry is updated quarterly and is derived from a search of more than 200,000 citations in the MEDLINE, EMBASE®, and CINAHL® databases, last updated prior to this report on June 14, 2003, August 6, 2002, and May 28, 2003, respectively. As of this writing, the registry contains approximately 2,500 studies, with another 3,000 studies pending full text assessment. The registry includes the full bibliographic reference (including MEDLINE index terms) and details the type of study, interventions considered, and targeted behavior. With the assistance of EPOC, the authors developed searches within the registry using the applicable clinical area MESH terms.

This approach differs from EPOC in one significant respect: it is EPOC policy to exclude interventions that do not involve provider or organizational change (e.g., patient education, self-management, and behavior change). In part because of this difference in scope, the authors conducted independent MEDLINE and hand searches for the first two priority conditions: diabetes and hypertension.

Inclusion and Exclusion Criteria

To begin, teams consisting of one or two senior reviewers (including an editor), trained two or more junior reviewers (junior faculty, fellows, and research assistants) to perform literature searches, conduct content reviews, and abstract data. The searches undertaken by these individuals revealed several thousand abstract titles for each priority condition. Stage 1 centered on the triage process for the article titles and/or abstracts (see Appendix C—triage forms), to determine if an article described an actual QI strategy. At this stage of the review process, randomized controlled trials, quasi-randomized trials, controlled before–after studies, interrupted time series, and before–after comparisons all were considered evaluations. A senior reviewer confirmed exclusion decisions using a random sample of 500 citations from the articles excluded at Stage 1–roughly a 20% sample. If the exclusion sample revealed any articles that should have been passed on for a full-text review, all the excluded citations were re-reviewed. The investigators included studies that examined the use of single or multiple QI strategies, with one exception: studies that used only patient education interventions were excluded because these studies likely will become the focus of a subsequent review. Studies were identified as relevant to quality improvement for this project if any one of the following applied:

1. The intervention was designed to increase the proportion of patients receiving recommended processes of care (e.g., those demonstrated to improve outcomes for patients with the condition of interest), including aspects of diagnosis and screening, therapeutic interventions, and patient education or counseling.

2. The intervention implemented organizational or structural features likely to benefit patients with the condition of interest.

3. The intervention attempted to improve outcomes for a broad and relatively unselected group of patients with the condition(s) of interest [e.g., “all patients with diabetes (or asthma, hypertension, etc.] who receive care at a clinic”).

4. The intervention targeted a subset of patients that is typically excluded from clinical research (i.e., frail elderly, minorities, homeless).

5. The intervention involved any of the specific QI strategies or sub-strategies noted in Table 1: provider reminder systems, facilitated relay, audit and feedback, provider educational interventions, patient educational interventions, promotion of self-management, patient reminders, organizational change, and financial, regulatory, or legislative incentives and interventions.

The authors set out to assess QI strategy effectiveness. The inclusion/exclusion criteria did not consider whether there was an established evidence-based guideline for the priority condition being studied. Nor did they review the evidence for the underlying quality improvement target. For example, the reviewers did not attempt to correlate the evidence for tight blood pressure

control with improved diabetes outcomes. Rather, they examined the evidence for QI interventions that have a positive effect on blood pressure control.

In Stage 2, a senior reviewer reconfirmed the description of a QI strategy in each included report and identified the study design (see Appendix C—triage forms). Determinations were made with regard to the study designs suitable for Stage 3 abstraction, based on the availability of the highest quality studies for that priority condition. Any study that was not excluded in Stage 2, on the basis of the title or abstract, was advanced to Stage 3 and a full text review.

In fact, the articles that remained part of the study at Stage 3 were scrutinized independently by two reviewers. Each reviewer abstracted information from the complete article about the QI strategy employed, the study design, and the outcomes evaluated (see Appendix D—Stage 3 abstraction forms). The forms used for the abstractions were tailored to each of the priority conditions, while still containing some common elements. Given the data available in the published literature, an emphasis was placed on information relevant to the effectiveness of the strategy and the aspects of study design most pertinent to the applicability of the study. The goals of health care quality (safety, effectiveness, patient-centeredness, timeliness, efficiency, and equity), outlined in the IOM's Quality Chasm report[4], also served to guide the reviewers. Unfortunately, most of these dimensions generally are not reported in studies assessing the efficacy of quality improvement strategies. Comparative data has been included, where available.

The purpose of Stage 3 was to ensure the exclusion of articles that were deemed to be something other than evaluations of QI strategies, and to allow an assessment of the amount and types of evidence available for a given priority condition. This information guided decisions regarding the breadth of the analysis to be undertaken, and how best to create discrete substrategies for synthesis. Any conflicts that arose in Stage 3 were resolved by consensus opinion between a junior reviewer and a senior reviewer.

Once the study designs and QI strategies were identified for a specific priority condition, articles meeting these final criteria underwent Stage 4 review. A junior reviewer conducted a detailed abstraction of relevant data (e.g., patient population, QI strategy, outcomes) from all included articles within the defined scope (see Appendix D—Stage 4 abstraction forms). A senior reviewer further confirmed the accuracy of the data abstraction.

Types of Evidence Assessed in the Review

The highest quality evidence available was used to assess the value of the QI strategies. Each of the study designs for the different QI strategies was assessed with respect to the conditions under consideration (i.e., hypertension). The reviewers also assessed important features of study conduct and analysis including concealment of allocation, patient blinding, provider blinding, and the unit of analysis relation to the unit of randomization. The hierarchy of study designs in Table 2 was used to guide the selection of study types for detailed data abstraction. Randomized controlled trials were considered the most persuasive source of evidence and so were deemed Level 1. If there were few or no randomized trials for a given strategy, the researchers evaluated Level 2 studies. Additionally, Level 2 studies also were reviewed to determine if findings about QI strategies were consistent across different study designs. Upon completing their initial literature review for each priority condition, the authors determined if sufficient studies existed meeting either Level 1 or Level 2 criteria. If they did, no detailed review of the study designs was performed at Level 3 (see Table 2). This is because biases commonly appear in Level 3

studies that make interpretation difficult, despite any insights they might provide with regard to applicability (e.g., external validity). Level 4 evidence was excluded; as such, no uncontrolled studies were considered. The studies also were catagorized by types of outcomes measured. Studies that did not report any of the outcome types specified in Table 3 also were excluded.

Table 2. Hierarchy of study designs

Level 1. Randomized controlled trials

Level 2. Controlled Before–After (CBA)—contemporaneous observation periods for control and intervention groups before and after an intervention

Interrupted time series (ITS)—well-defined time period for intervention implementation and at least three time points both before and after intervention

Quasi-randomized trials—contained at least two cohorts of patients assembled prospectively based on an allocation procedure that was non-random, but arbitrary, in the sense of bearing no apparent connection to patient or provider factors that might affect intervention outcome (e.g. alternation, date of birth, even/odd character of provider or patient identification)

Level 3. Observational studies with controls—includes before-after and time series not meeting strict definitions of CBA and ITS (above), case–control studies, cohort studies with controls.

Level 4. Observational studies without controls (e.g., cohort studies without controls and case series)

Table 3. Outcomes relevant to inclusion criteria*

	Measures of Disease Identification	Measures of Disease Control	Measures of Provider Adherence	Measures of Patient Adherence
Included studies reported one of more of the outcomes presented here	Proportion of eligible patients receiving appropriate screening (i.e., blood pressure measurements, cancer screening)	Clinical outcomes Intermediate outcomes with established connections to clinical outcomes (e.g., HbA_{1c}, blood pressure, lipid levels)	Performance of specific processes of care with established connections to patient outcomes Adherence to well-recognized practice guidelines (e.g., from professional societies)	Biochemical assays (e.g., blood or urine drug levels, urine cotinine for smoking cessation) Pharmacy data (e.g., refilled prescriptions) Home or office pill counts Patient interviews

*Measures of provider knowledge, patient understanding, self-efficacy, or other intermediate outcomes were included only when they accompanied outcomes listed in the above table. For instance, a study reporting a measure of patient adherence with care as well as changes in patient understanding, self-efficacy, or empowerment would be included. Similarly, articles reporting only measures of satisfaction with care or resource use were not included unless they also reported measures of disease identification, disease control, or provider or patient adherence.

Evaluation of Quality Improvement Strategies

Most of the reported information addressed QI strategy effectiveness. There was a paucity of available data on the safety, equity, and applicability of the various approaches.

A number of factors may influence the success of a QI strategy. Table 4 summarizes many of these factors and organizes them into three categories. Relatively little information on the features of the QI target was obtainable, due to time restrictions. This is a potential limitation of the analysis. The authors have noted in the table those factors for which information was obtained, as well as those factors included in the synthesis.

Table 4. Features that may affect success of QI interventions

Features of the Study	Features of the QI Intervention	Features of the QI Target
Study setting √ Study period √ Country (√√ US vs. non-US) __ Financial/organizational structure of health care system √ Type of clinical setting (e.g., general vs. specialty clinic, community-based, work site intervention) √ patient population (e.g., specific disease being studied (diabetes vs. hypertension vs. asthma, etc.), early vs. advanced illness, significant comorbid conditions, underserved, poor adherence) √ magnitude of local quality gap *Study methodology* √√ Trial design (e.g., RCT, quasi-RCT, CBA, ITS)† √√ Concealment of allocation √√ Blinding (patients, providers) √√ Agreement in unit of randomization and unit of analysis	*Type of QI strategy* √√ Broad category of QI strategy (e.g., patient education, provide education, audit & feedback, etc.) √√ Number of QI strategies employed in the intervention * Intensity of QI strategy (e.g., number of educational sessions, frequency of audit & feedback cycles, extent of case management) __ Involvement of top management and other forms of institutional support * Format in which QI strategy delivered (e.g., face-to-face, dissemination of printed materials) √√ Use of an information system	*Content* __ Attitude of clinicians toward target (driven in turn by guideline complexity, evidence base, concordance with existing practice) __ Complexity of action required by provider (e.g., making a referral, ordering a test, adjustment of medication regimen, performing specific aspects of history or physical) __ Baseline level of adherence with target __ Difficulty in achieving target (e.g., achieving a specific goal such as blood pressure below a certain value vs. process performance irrespective of outcome)

† RCT = randomized controlled trial; CBA = controlled before–after study; ITS = interrupted time series.
__ Indicates no information collected
* Indicates information captured in text answers by reviewers rather than structured format
√ Indicates data collected relevant to this feature
√√ Indicates data collected and included in summary analysis, when feasible for a given topic

Quantitative Synthesis of Quality Improvement Strategies

Quantitative evaluations of the QI effect were performed for the various strategies, when possible. These evaluations were done only in situations when: 1) a sufficient number of studies with similar outcomes were available, and 2) the studies were sufficiently homogeneous in their design and population to provide a valid quantitative sample.

Calculation of summary effect for studies. In addition to the descriptive and qualitative investigations, two additional forms of analysis were planned for inclusion in the review. The first involved calculation of the median effect for outcomes within a given category (i.e., all provider adherence outcomes reported by a given study) so that studies with the same features could be compared using a common metric. Following the method employed in a recent systematic review of strategies for guideline implementation,[61] researchers identified for each study the adherence outcome that indicated the median improvement attributable to the intervention. For example, if a study reported one outcome involving adherence to a guideline for checking HbA_{1c}, another relating to managing cardiovascular risk factors, and another for delivery of patient education, a calculation of the net improvement attributable to the intervention for each outcome then would be done. The net improvements then were ranked for all of the outcomes and the median net improvement was used as a summary measure for the study.

The net improvement in adherence was calculated as (Post-intervention adherence – Pre-intervention adherence)$_{\text{Study group}}$ – (Post-intervention adherence–Pre-intervention adherence)$_{\text{Control group}}$. Outcomes were not combined for measures of disease control, so for example, the authors simply reported the net reduction in HbA_{1c}, systolic blood pressure (SBP) or diastolic blood pressure (DBP) attributable to the intervention.

For instance, the net reduction in SBP attributable to the intervention was calculated as:

(Post-intervention SBP – Pre-intervention SBP)$_{\text{Study group}}$ – (Post-intervention SBP – Pre-intervention SBP)$_{\text{Control group.}}$

To characterize the impact of a particular type of QI strategy (i.e., provider education) or study feature (i.e., trial design), a calculation was made of the median effect achieved in studies with the feature of interest. For instance, all trials with interventions that included some aspect of provider education and also reported a change in mean SBP for the study groups were identified. Next, the median net reduction in SBP for these trials was computed and compared to the median effect for all trials, as well as the median effect for trials with interventions having no component of provider education. The median improvement in adherence across different QI types was compared similarly.

The use of median effects, rather than average effects, prevented the skewing of summary measures based on outliers with particularly large or small effect sizes. This was regarded as particularly important because, if publication bias were present, small studies with relatively large effect sizes would more likely be published than small studies with more modest effect size. Thought was given to a weighted median, with weights based on sample size, to avoid giving equal weight to all studies regardless of size. Weighted medians are not as straightforward as weighted means, especially when attempting to preserve the original significance of the effect size (e.g., the observed reduction in HbA_{1c} or SBP in the units used for those outcomes). So

rather than attempting a weighting function, the authors chose instead to examine the median effect sizes using different strata of study sample size (e.g., comparing the median effect among studies with sample sizes in the lowest quartile vs. those in the highest quartile, or lower half vs. upper half).

Adjustments for unit of analysis errors. The "clustering effect," in which the unit of analysis and unit of allocation differ (i.e., providers or clinics randomized, but patient level outcomes analyzed) was anticipated in a significant number of studies. The significance of clustering is that patients within a cluster are not independent (e.g., patients at one clinic resemble one another in more ways than they resemble patients at other sites, or those cared for by other providers in the trial). Unit-of-analysis errors do not affect point estimates for effect sizes, but they can have a spurious narrowing effect on the associated confidence interval, causing potentially false-positive trial results.[62-69] To prevent the same false precision in this analysis, an effective sample size[*] was calculated for each study for the meta-regressions described below. Moreover, the degree to which investigators acknowledged or accounted for cluster effects did not affect the analysis, apart from the fact that investigators who did consider cluster effects in the design or analysis of their trial were more likely to report data such as the number of providers randomized, rather than reporting only the total numbers of patients in each group. The same investigators also might provide more technical details, such as values for the intra-cluster coefficient (ICC).[70-72]

Meta-regression Analyses

For the more involved quantitative analyses—meta-regression analysis of included studies—the investigators used a more conventional measure of effect size, defined as the difference between the mean values for the intervention and control arms, divided by the pooled estimate of groups within the standard deviation.[†] The researchers constructed these formal effect sizes, as well as the above median effect measures, such that a positive result always reflected improvement (e.g., a positive reduction in average HbA_{1c} or a positive improvement in adherence).

The regression models aimed to evaluate the relative effectiveness of different intervention components and the impact of study features such as trial design and study period. Specifically, the investigators constructed regression models using the pre-intervention effect size (ES_{Pre}) as a predictor variable. Initially, each methodological feature or QI strategy was modeled with ES_{Pre} to evaluate its effect on the post-intervention effect size (ES_{Post}); subsequently, the researchers developed multivariate models using multiple components as an individual feature's covariates, in order to independently assess the effect of an individual feature after adjustment for other components. Linear regression was carried out as $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$, with $X_1 = ES_{pre}$ and the dependent variable, Y, corresponding to the outcome of interest—a measure of disease control such as HbA_{1c}, or the summary measure of adherence outcome described above. The approach retained ES_{pre} as a predictor in all analyses because baseline differences between the study and

---

[*] Effective N = (km) / (1 + (m-1)r) where k is the number of clusters and m is the number of observations per cluster and r is the intra-cluster coefficient. When r = 0, then N = km. When r = 1, then N = k

[†] Effect size = $(\bar{X}_I - \bar{X}_C) / S_p$ where $\bar{X}_I$ is the mean for the intervention group, $\bar{X}_C$ is the mean for the control group and S_p is the pooled-within-groups standard deviation, which is calculated from: $S_p^2 = ((N_I - 1) S_I^2 + (N_C - 1) S_C^2) / (N_I + N_C - 2)$. N_I and N_C are the intervention and control sample sizes and S_I and S_C are the intervention and control standard deviations.

control groups were expected to act as important covariates, even when these differences did not meet conventional thresholds for statistical significance.