The Role of Formative Evaluation in Implementation Research and the QUERI Experience

Cheryl B. Stetler, PhD, RN, FAAN,[1] Marcia W. Legro, PhD,[2,3] Carolyn M. Wallace, PhD,[2] Candice Bowman, PhD, RN,[4] Marylou Guihan, PhD,[5] Hildi Hagedorn, PhD,[6] Barbara Kimmel, MS, MSc,[7,8] Nancy D. Sharp, PhD,[2] Jeffrey L. Smith (PhD candidate)[9]

[1]Independent Consultant, Amherst, MA, USA; [2]VA Puget Sound Health Care System, Seattle, WA, USA; [3]University of Washington, Seattle, WA, USA; [4]VA San Diego Healthcare System, San Diego, CA, USA; [5]Edward Hines, Jr. VA Healthcare System, Hines, IL, USA; [6]Minneapolis VA Medical Center, Minneapolis, MN, USA; [7]Baylor College of Medicine, Houston, TX, USA; [8]Houston Veterans Affairs Medical Center, Houstan, TX, USA; [9]Central Arkansas Veterans Healthcare System, Little Rock, AR, USA.

This article describes the importance and role of 4 stages of formative evaluation in our growing understanding of how to implement research findings into practice in order to improve the quality of clinical care. It reviews limitations of traditional approaches to implementation research and presents a rationale for new thinking and use of new methods. Developmental, implementation-focused, progress-focused, and interpretive evaluations are then defined and illustrated with examples from Veterans Health Administration Quality Enhancement Research Initiative projects. This article also provides methodologic details and highlights challenges encountered in actualizing formative evaluation within implementation research.

A s health care systems struggle to provide care based on well-founded evidence, there is increasing recognition of the inherent complexity of implementing research into practice. Health care managers and decision makers find they need a better understanding of what it takes to achieve successful implementation, and they look to health care researchers to provide this information. Researchers in turn need to fill this need through collection of new, diverse sets of data to enhance understanding and management of the complex process of implementation.

A measurement approach capable of providing critical information about implementation is formative evaluation (FE). Formative evaluation, used in other social sciences, is herein defined as *a rigorous assessment process designed to identify potential and actual influences on the progress and effectiveness of implementation efforts*. Formative evaluation enables researchers to explicitly study the complexity of implementation projects and suggests ways to answer questions about context, adaptations, and response to change.

The Department of Veterans Affairs (VA) Quality Enhancement Research Initiative (QUERI) has integrated FE into its implementation program.[1–3] This article introduces QUERI and its implementation focus. It then describes research challenges that call for the use of FE in this specialized field of study, reviews FE relative to QUERI implementation research, identifies 4 evaluative stages, and presents challenges to the conduct of FE.

THE VETERAN HEALTH ADMINISTRATION'S QUERI PROGRAM

The Quality Enhancement Research Initiative, begun in 1998, is a comprehensive, data-driven, outcomes-based, and output-oriented improvement initiative.[2,3] It focuses on identification and implementation of empirically based practices for high-risk/high-volume conditions among the veteran population and on the *evaluation* and *refinement* of these implementation efforts.[3] The Quality Enhancement Research Initiative's innovative approach[1–4] calls upon researchers to work toward rapid, significant improvements through the systematic application of best clinical practices. It also calls upon researchers to study the implementation process to enhance and continuously refine these quality improvement (QI) efforts.[1–4]

Classic intervention research methods[5,6] provide the means to evaluate targeted outcomes of implementation/QI efforts. From an evaluation perspective, studies using intervention designs, such as a cluster-randomized trial or quasi-experimental approaches, routinely include a *summative evaluation*. Summative evaluation is a systematic process of collecting data on the impacts, outputs, products, or outcomes hypothesized in a study.[7] Resulting data provide information on the degree of success, effectiveness, or goal achievement of an implementation program.

In an action-oriented improvement program, such as QUERI, summative data are essential but insufficient to meet the needs of implementation/QI researchers. Evaluative information is needed beyond clinical impact of the change effort and beyond discovering whether a chosen adoption strategy worked. Implementation researchers need to answer critical questions about the feasibility of implementation strategies, degree of real-time implementation, status and potential influence of contextual factors, response of project participants, and any adaptations necessary to achieve optimal change. Formative evaluation provides techniques for obtaining such information and for overcoming limitations identified in early implementation/QI studies.

Table 1. A Spectrum of Definitions of Formative Evaluation

"Evaluative activities undertaken during the design and pretesting of programs to guide the design process"[17]

" . . . a method of judging the worth of a program while the program activities are forming or happening. Formative evaluation focuses on the process"[18]

An assessment that focuses on "the internal dynamics and actual operations of a program in order to understand its strengths and weaknesses and changes that occur in it over time"[19]

NEED FOR FE IN IMPLEMENTATION/QI RESEARCH

The RE-AIM framework of Glasgow and colleagues highlights critical information that is missing from current research publications—i.e., information needed to evaluate a study's *potential for translation and public health impact*.[8,9] Such information includes the *efficacy/effectiveness* of an intervention, its *reach* relative to actual/representative subject participation rate, its *adoption* relative to actual/representative setting participation rate, its *implementation* or intervention *fidelity*, and its *maintenance* over time.

The focus of the RE-AIM framework is the study of health promotion interventions. Similar issues must be addressed during implementation research if potential adopters are to replicate critical implementation processes. In addition, implementation researchers need to capture in-depth information on participant and contextual factors that facilitate or hinder successful implementation. Such factors can be used during the project to optimize implementation and inform post hoc interpretation.

As implementation efforts can be a relatively messy and complex process, traditional study designs alone are often inadequate to the task of obtaining evaluative information. For example, randomized clinical trials (RCT) may leave questions important to system-wide uptake of targeted research unanswered. As Stead et al.[10,11] suggest, traditional intervention research can fail to "capture the detail and complexity of intervention inputs and tactics" (10, p. 354), thereby missing the true nature of interventions as well as significant organizational factors important for replication.[10,11]

Another argument for performing FE has been highlighted in guideline/QI literature, i.e., the need to address potential interpretive weaknesses. Such weaknesses relate to a failure to account for key elements of the implementation process and may lead to unexplainable and/or poor results. For example, Ovretveit and Gustafson[12] identified *implementation assessment failure, explanation failure,* and *outcome attribution failure. Implementation assessment failure* can lead to a "Type III" error, where erroneous study interpretations occur because the intervention was not implemented as planned.[12,13] *Explanation* and *outcome attribution* relate to failures to explore the *black box* of implementation. Specifically, what actually did/did not happen within the study relative to the implementation plan, and what factors in the implementation setting, anticipated or unanticipated, influenced the actual degree of implementation? By failing to collect such data, potential study users have little understanding of a particular implementation strategy. For example, 1 study regarding opinion leadership did not report the concurrent implementation of standing orders.[14]

Use of a traditional intervention design does not obviate collection of the critical information cited above. Rather, complementary use of FE within an experimental study can create a dual or hybrid style approach for implementation research.[15] The experimental design is thus combined with descriptive or observational research that employs a mix of qualitative and quantitative techniques, creating a richer dataset for interpreting study results.

FORMATIVE EVALUATION WITHIN QUERI

As with many methodologic concepts, there is no single definition/approach to FE. In fact, as Dehar et al.[16] stated, there is a decided "lack of clarity and some disagreement among evaluation authors as to the meaning and scope" of related concepts (16, p. 204; see Table 1 for a sampling). Variations include differences in terminology, e.g., an author may refer to *FE, process evaluation,* or *formative research*.[16,17]

Given a mission to make rapid, evidence-based improvements to achieve better health outcomes, the authors have defined FE as *a rigorous assessment process designed to identify potential and actual influences on the progress and effectiveness of implementation efforts.* Related data collection occurs before, during, and after implementation to optimize the potential for success and to better understand the nature of the initiative, need for refinements, and the worth of extending the

Table 2. Potential Uses of Formative Evaluation[10,13,16,20–27]

Understand the nature of the local implementation setting

Assess whether a program or intervention addresses a significant need

Modify a proposed program or intervention, as needed

Determine the extent, fidelity, and qualities of the implementation of an intervention program . . . (e.g., to) describe the activities actually implemented. . . . (and) . . . explain program operations[21]

Systematically detect and monitor unanticipated events (and adjust if appropriate)

Optimize/control implementation to improve the potential for success

Obtain ongoing input for short-term adjustments

Document continual progress

Inform future similar implementation efforts, e.g., within other health care sites or a larger system

Avoid type III errors: "Failing to detect differences between the original intervention plan and the ultimate manner of implementation"[13]; or failure to understand how complex the phenomena of interest really are

Understand the extent/dose, consistency, usefulness, context, and quality of an intervention's implementation

Understand the nature and implications of local adaptation

Assist interpretation of program outcomes or worth in terms of the effort required to achieve a designated level of improvement

Foster an understanding of the causal events leading to change and the specific components of the intervention that most influenced it[20]

Standardize on-going implementation

Understand the experience of those directly affected by implementation efforts

Table 3. An Example of Developmental FE

Spinal cord injury (SCI) QUERI: respiratory vaccine initiative

Implementation goal: one evidence-based strategy to improve preventive care delivery rates is a practitioner reminder system. As the VHA had a nationally developed and distributed computerized clinical reminder (CCR) for influenza that did not appear to be in wide use at study sites, SCI QUERI set a goal of having the CCR installed locally and used by practitioners in the SCI centers across the country.

FE activity: electronic questionnaires with follow-up interviews about the influenza CCR were administered to computer application coordinators at VA facilities with SCI centers and to personnel at each SCI center. From these data it was clear that the current version of the reminder was installed at all locations, but that local methods of installing and using the CCR varied widely. This indicated that the initial plan to develop a standard set of instructions was inappropriate. As an alternative, the research team developed general principles and encouraged SCI personnel to seek further help from their local information technology staff.[35,36]

Value: the information from this activity guided the research team's choice of methods to assist SCI center personnel to move toward the goal of using the CCR.

QUERI, quality enhancement research initiative; SCI, spinal cord injury; VHA, Veterans Health Administration; CCR, computerized clinical reminder; FE, formative evaluation.

project to other settings. This approach to FE incorporates aspects of the last 2 definitions in Table 1 and concurs with the view that *formative* connotes action.[16] In QUERI, this action focus differentiates FE from "process" evaluations where data are not intended for concurrent use.

Various uses of FE for implementation research are listed in Table 2. Uses span the timeframe or stages of a project, i.e., development/diagnosis, implementation, progress, and interpretation. Within QUERI, these stages are progressive, integrated components of a single hybrid project. Each stage is described below, in the context of a single project, and illustrated by QUERI examples (Tables 3, 4, and 6–10). Each table provides an example of 1 or more FE stages. However, as indicated in some of the examples, various evaluative activities can serve multiple stages, which then merge in practice. Formative evaluation at any stage requires distinct plans for adequate measurement and analysis.

Developmental Evaluation

Developmental evaluation occurs during the first stage of a project and is termed a *diagnostic analysis*.[1,28] It is focused on enhancing the likelihood of success in the particular setting/s of a project, and involves collection of data on 4 potential influences: (a) actual degree of less-than-best practice; (b) determinants of current practice; (c) potential barriers and facilitators to practice change and to implementation of the adoption strategy; and (d) strategy feasibility, including perceived utility of the project. (Note:

studies conducted to obtain generic diagnostic information prior to development of an implementation study are considered formative research, not FE. Even if available, a diagnostic analysis is suggested given the likelihood that generically identified factors will vary across implementation sites.)

Activity at this stage may involve assessment of known prerequisites or other factors related to the targeted uptake of evidence, e.g., perceptions regarding the evidence, attributes of the proposed innovation, and/or administrative commitment.[11,21,29–31] Examples of formative diagnostic tools used within QUERI projects include organizational readiness and attitude/belief surveys[32,33] (also see Tables 3 and 7). Such developmental data enable researchers to understand potential problems and, where possible, overcome them prior to initiation of interventions in study sites.

In addition to information available from existent databases about current practice or setting characteristics, formative data can be collected from experts and representative clinicians/administrators. For example, negative unintended consequences might be prospectively identified by key informant or focus group interviews. This participatory approach may also facilitate commitment among targeted users.[34]

Implementation-Focused Evaluation

This type of FE occurs throughout implementation of the project plan. It focuses on analysis of discrepancies between the plan

Table 4. Implementation-Focused FE

Congestive heart failure (CHF) QUERI

Implementation goals: four clinical strategies were identified from research to improve health status and reduce readmission rate costs for CHF patients, i.e., identifying patients, determining readiness for discharge, patient education, and rigorous outpatient follow-up.[41,42] One evidence-based recommendation within the patient education component was weight management. To self-manage weight, patients were asked to weigh themselves every day.

FE activity: following development of individualized implementation intervention plans and timelines for each site, CHF QUERI staff conducted weekly teleconferences with key local contacts. Using structured interview questions, barriers and facilitators were identified and potential strategies to resolve the barriers were discussed. The following week, sites were provided with any modification of procedures or forms (mid-course corrections). One barrier to accomplishing the weight management goal was uncovered during weekly calls. The Prosthetic Department at 1 site did not interpret current policy in a manner that allowed provision of scales for CHF patients. The research team of CHF QUERI was able to negotiate a policy change, so access to a scale was no longer a barrier to self-monitoring fluid retention when patients were discharged.

Value: the weekly teleconferencing and problem resolutions changed the perceptions of clinical staff that this research was not related to day-to-day activities. The calls also provided a model of problem-solving behaviors for the clinical staff (e.g., Plan, Do, Study, Act) as well as information for the research team that enabled "user friendly" modifications of original project tools. Overall, changes at the system level improved the probability that patient education would lead to better self-monitoring of weight.

QUERI, quality enhancement research initiative; FE, formative evaluation; CHF, congestive heart failure.

Table 5. Critical Measures of Implementation

1. *Integrity of the innovation*[43–45]

   Data are needed that specify the *fidelity* of the implementation intervention, i.e., the extent to which the intervention was actually implemented or *delivered*. For example, to what extent did the opinion leader "lead"; in what way and to what degree did administration provide "support"; or to what extent did a case manager or educational outreach worker perform role expectations?

   Clear operational definitions for each component of the implementation intervention are required. These should relate to the conceptual rationale for the intervention's selection.

   When variable degrees of delivery are possible, the *dose* of the intervention *delivery* should be measured. If relevant, this quantitative score may be correlated with targeted outcomes.

2. *Exposure to the innovation*

   Data are needed that specify the degree to which delivered products are actually *experienced* by the targeted users. For example, education may be 1 element of an implementation strategy that is delivered, yet may only reach or be accessed by a limited number of targeted users; likewise, written or web-based materials may be disseminated but not received or read.

   Clear operational definitions for each related exposure are required. These too should relate to the conceptual rationale for the intervention's selection.

   When variable degrees of experience are possible, the *dose* of the exposure should be measured. This quantitative score can be correlated with targeted outcomes. Also, it is possible that an intermediate result of exposure should be measured, i.e., a change in attitude or knowledge, or engagement in a targeted process, rather than mere presence/attention.

3. *Intensity of implementation*

   When multifaceted interventions are used, an overall *implementation or intensity* score can be considered. This overall level of effort may relate to the number of change interventions. However, it is probably more productive to measure intensity both within and across conceptual categories of interventions, e.g., efforts geared to change knowledge versus behavior versus systems.

   An overall score per site can be calculated as Boyd and Windsor[46] did with their numeric "Program Implementation Index." Another approach is use of goal attainment scaling.[47,48] This enables comparison across sites or units within a site, especially in those studies where "local adaptation" or choice of alternative interventions is a selected strategy.

---

and its operationalization and identifies influences that may not have been anticipated through developmental activity. As Hulscher et al. note in a relevant overview of "process" evaluation, FE allows "researchers and implementers to (a) describe the intervention in detail, (b) evaluate and measure actual exposure to the intervention, and (c) describe the experience of those exposed (13, p. 40)" — concurrently. It also focuses on the dynamic context within which change is taking place, an increasingly recognized element of implementation.[37–40]

Implementation-focused formative data enable researchers to describe and understand more fully the major barriers to goal achievement and what it actually takes to achieve change, including the timing of project activities. By describing the actuality of implementation, new interventions may be revealed. In terms of timing, formative data can clarify the true length of time needed to complete an intervention, as failure to achieve results could relate to insufficient intervention time.

Implementation-focused formative data also are used to keep the strategies on track and as a result optimize the likelihood of affecting change by resolving actionable barriers, enhancing identified levers of change, and refining components of the implementation interventions. Rather than identify such modifiable components on a post hoc basis, FE provides timely feedback to lessen the likelihood of type III errors (see Tables 4, 6, 7, and 9).

In summary, FE data collected at this stage offer several advantages. They can (a) highlight actual versus planned interventions, (b) enable implementation through identification of modifiable barriers, (c) facilitate any needed refinements in the original implementation intervention, (d) enhance interpretation of project results, and (e) identify critical details and guidance necessary for replication of results in other clinical settings.

Measurement within this stage can be a simple or complex task. Table 5 describes several critical issues that researchers should consider. As with other aspects of FE, both quantitative and qualitative approaches can be used.

Progress-Focused Evaluation

This type of FE occurs during implementation of study strategies, but focuses on monitoring impacts and indicators of

---

Table 6. Implementation and Progress-Focused FE

Substance use disorders QUERI

Implementation goal: improve opioid agonist clinics' implementation of 4 best-practice recommendations: (1) adequate dosing, (2) adequate counseling support, (3) maintenance orientation, and (4) contingency management interventions.

FE activity[50,51]:

i. Prototype toolkits for implementing QI and monitoring clinic practices and patient outcomes were distributed to 8 pilot clinics. A minimum of monthly contact between implementation team staff and clinic staff assessed barriers, strategies to overcome barriers, suggestions for improving tools or adding additional tools, and progress. Toolkits were modified to meet the needs of the clinics. Satisfaction with the modifications was evaluated on the next scheduled call.

ii. Pilot clinics also submitted monthly data on practice variables (e.g., mean dose, counseling frequency) and patient outcomes (e.g., percent of urine screens positive for illicit substances).

Value:

i. The original toolkit was modified to improve its usefulness. The exchange of information and responsiveness of the implementation team to clinic suggestions improved working relationships.

ii. Data were used for QI goal setting and modification as initial goals were achieved. Data were also used to monitor progress and provide feedback to sites. Formative evaluation allowed quick identification of stagnation of progress that suggested the need for increased facilitation or introduction of new tools to sustain enthusiasm for the intervention.

QUERI, quality enhancement research initiative; FE, formative evaluation; QI, quality improvement.

Table 7. Developmental/Implementation/Progress FE

SCI QUERI: respiratory vaccine initiative

Implementation goal: research evidence indicates that persons who have negative attitudes about influenza vaccine (safety, effectiveness, utility) are less likely to seek it out annually. So, SCI QUERI identified both providers and patients to receive educational information to directly and indirectly reduce such negative attitudes.

FE activity: attitude data were collected from both SCI staff[52] and patients[53] longitudinally over 3 years by questionnaire to both identify initial issues with beliefs about vaccination and to track changes over time.

Value: data regarding prevalent negative attitudes were used to construct targeted questionnaires as well as educational materials to counter these attitudes. For example, staff materials during the second year were changed to include current data regarding their patients in terms of the percent who endorsed specific incorrect statements about vaccines.

QUERI, quality enhancement research initiative; FE, formative evaluation; SCI, spinal cord injury.

progress toward goals. The proactive nature of FE is emphasized, as progress data become feedback about the degree of movement toward desired outcomes. Using implementation data on dose, intensity, and barriers, factors blocking progress may be identified. Steps can then be taken to optimize the intervention and/or reinforce progress via positive feedback to key players. As Krumholz and Herrin[49] suggest, waiting until implementation is completed to assess results "obscures potentially important information … about trends in practice during the study [that] could demonstrate if an effort is gaining momentum—or that it is not sustainable" (see Tables 6 and 7).

Interpretive Evaluation

This stage is usually not considered a type of FE but deserves separate attention, given its role in the illumination of the black box of implementation/change. Specifically, FE data provide alternative explanations for results, help to clarify the meaning of success in implementation, and enhance understanding of an implementation strategy's impact or "worth." Such "black box" interpretation occurs through the end point triangulation of qualitative and quantitative FE data, including associational relationships with impacts.

Interpretive FE uses the results of all other FE stages. In addition, interpretive information can be collected at the end of the project about key stakeholder experiences. Stakeholders include individuals expected to put evidence into practice as well as those individuals expected to support that effort. These individuals can be asked about their perceptions of the implementation program, its interventions, and changes required of them and their colleagues.[10,13,27,38,46,54] Information can be obtained on stakeholder views regarding (a) usefulness or value of each intervention, (b) satisfaction or dissatisfaction with various aspects of the process, (c) reasons for their own program-related action or inaction, (d) additional barriers and facilitators, and (e) recommendations for further refinements.

Information can also be obtained regarding the degree to which stakeholders believe the implementation project was successful, as well as the overall "worth" of the implementation effort. Statistical significance will be calculated using the summative data. However, as inferential statistical significance does not necessarily equate with clinical significance, it is useful to obtain perceptions of stakeholders relative to the "meaning" of statistical findings. For some stakeholders, this meaning will be placed in the context of the cost of obtaining the change relative to its perceived benefits (see Tables 8–10).

Formative evaluation, as a descriptive assessment activity, does not per se test hypotheses. However, within an experimental study, in-depth data from a concurrent FE can provide working hypotheses to explain successes or failures, particularly when the implementation and evaluation plans are grounded in a conceptual framework.[55–57] In this respect, interpretive FE may be considered as case study data that

Table 8. Interpretive FE

Mental health (MH) QUERI

Implementation goal: an antipsychotic treatment improvement program (ATIP) was designed to implement an evidence-based treatment model for patients with schizophrenia. A second program, TIDES/WAVES, was designed to improve care for patients with depression.

FE activity: for both projects a series of formative evaluations continuously assessed the extent and nature of facilitation by the research team, intervention activities conducted by site personnel, and related barriers and facilitators.

Value:

i. In the ATIP project, FE demonstrated a correlation between the amount of external facilitation and improvement in antipsychotic dosing, the extent to which lack of clinician time inhibits dissemination and QI activities, and the acceptability and utility of MH QUERI tools/products.

ii. In the depression project, FE has generated information about factors influencing caseload capacity for nurse "care managers" who maintain regular contact with depressed patients to monitor treatment adherence, assess symptomatic improvement, provide patient education, and facilitate communication among patients, primary care clinicians, and mental health specialists.

Substance use disorder QUERI

Implementation goal: the opioid agonist therapy effectiveness initiative was designed to implement evidence-based treatment for patients with opioid dependence.[50,51]

FE activity: during the project, the research team took notes on all facilitation-related telephone calls. They also collected monthly, semi-structured interview data from program leaders. As the project moved forward, it was clear that some sites were very successful in applying the treatment plan while others stagnated.

Value: while the FE provided some suggestions about how to modify the intervention and tools, the data were also useful to accurately "tell the story" of the individual clinics and to illustrate differences between clinics that were successful with implementation versus those that were not.

QUERI, quality enhancement research initiative; FE, formative evaluation; QI, quality improvement; NH, mental health; ATIP, antipsychotic treatment improvement program.

Table 9. Implementation/Interpretive FE

HIV/AIDS QUERI

Implementation goal: as in many conditions, HIV care processes fall short of best practice recommendations. We sought to implement and evaluate real-time computerized clinical reminders (CR) and a collaborative intensive quality improvement program, based on the Institute for healthcare improvement breakthrough series (IQS).

FE activity:

i. By conducting ethnographic interviews and observations at the participating CR sites, we assessed existing organizational barriers and facilitators regarding use of reminders.

ii. The degree to which each site implemented the IQS improvement technique was measured at regular intervals using a Site Activation Scale (SAS) developed specifically for this project. Longitudinal tracking of SAS scores allowed calculation of time-to-improvement as well as qualitative comparisons of activation. We also assessed whether specific barriers or characteristics were associated with how well a site scored on activation.

Value:

i. We identified 6 barriers to the effective use of the CR intervention, including workload, lack of time to follow-up, inapplicability, limited training, interruption of patient-provider face time, and use of paper for physician orders at some sites. Seventeen prioritized short- and long-term recommendations were generated to improve the usefulness and usability of the 9 reminders.[59]

ii. By using the SAS scoring system, we were able to identify that higher performing IQS sites rapidly adopted and applied basic quality improvement concepts, like the PDSA (PLAN, Do, Study, Act). In contrast, we were able to see that lower performing IQS sites were slow to begin applying key concepts and had not completed an entire PDSA cycle well into an action period. Although the contrast between levels of activation was not always marked, the sites showing the most rapid adoption and generalization of quality improvement concepts tended to show significant improvements in care. IQS FE was also used to refine the intervention, to some extent, during the project, and to a major extent for the next "nationwide rollout" phase.

QUERI, quality enhancement research initiative; FE, formative evaluation; CR, clinical reminder; SAS, site activation scale.

---

contribute to theory building.[58] Overall, FE data may provide evidence regarding targeted components of a conceptual framework, insights into the determinants of behavior or system change, and hypotheses for future testing.

CHALLENGES OF CONDUCTING FE

Formative evaluation is a new concept as applied to health services research and as such presents multiple challenges. Some researchers may need help in understanding how FE can be incorporated into a study design. Formative evaluation is also a time-consuming activity and project leaders may need to be convinced of its utility before committing study resources. In addition, much is yet to be learned about effective approaches to the following types of issues:

1. In the well-controlled RCT, researchers do not typically modify an experimental intervention once approved. However, in improvement-oriented research, critical problems that prevent an optimal test of the planned implementation can be identified and resolved. Such actions may result in alterations to the original plan. The challenge for the researcher is to identify that point at which modifications create a different intervention or add an additional intervention. Likewise, when the researcher builds in "local adaptation," the challenge is to determine its limits or clarify statistical methods available to control for the differences. An implementation framework and clear identification of the underlying conceptual nature of each intervention can facilitate this process. As Hawe et al.[43] suggest, the researcher has to think carefully about the "essence of the intervention" in order to understand the actual nature of implementation and the significance of formative modifications.

2. Implementation and QI researchers may encounter the erroneous view that FE involves only qualitative research or that it is not rigorous, e.g., that it consists of "just talking to a few people". However, FE does not lack rigor nor is it simply a matter of qualitative research or a specific qualitative methodology. Rather, FE involves selecting among rigorous qualitative and quantitative methods to accomplish a specific set of aims, with a plan designed to produce credible data relative to explicit formative questions.[61]

3. A critical challenge for measurement planning is selection or development of methods that yield quantitative data for the following types of issues: (a) assessment of associations between outcome findings and the intensity, dose, or exposure to interventions and (b) measurement of the adapta-

---

Table 10. An Illustrative, Potential FE

Ischemic heart disease QUERI

Implementation goal: pilot intervention teams from 8 hospitals developed and implemented strategies to increase measurement and management of low-density lipoprotein cholesterol (LDL-c) in coronary heart disease patients. Interventions included audit/feedback, patient education, pharmacist case management, lipid clinics, order templates, and paper point-of-care reminders.

FE activity (an interpretive post hoc evaluation or missed FE opportunity[60]): researchers conducted structured interviews with participants to identify barriers and facilitators to implementation of pilot interventions. Interviews specifically addressed awareness of and agreement with secondary prevention guidelines, priorities for intervention activities during the patient visit, availability of laboratory and pharmacy data, and effectiveness of intervention planning and implementation. Results were organized and interpreted using theory-based content analysis based on the Promoting Action Research in Health Systems (PARIHS) framework, which identifies 3 components of successful implementation of evidence-based practice: evidence, context, and facilitation.[9,38]

Value: the analysis identified barriers to successful implementation that related primarily to the intervention process and secondarily to characteristics of the intervention context. Interview responses indicated that planning, including identification of resources and potential barriers and facilitators, was a critical and universally underutilized step in the intervention process. Data from the interviews suggested tools and guidelines to improve planning and implementation skills.

QUERI, quality enhancement research initiative; FE, formative evaluation.

tions of a "standard" protocol across diverse implementation settings.[62] Whether flexibility is a planned or unplanned component of a study, it should be measured in some consistent, quantifiable fashion that enables cross-site comparisons. Goal attainment scaling is 1 possibility.[47,48]

4. A final issue facing implementation researchers is how to determine the degree to which FE activities influence the results of an implementation project. If FE itself is an explicit intervention, it will need to be incorporated into recommendations for others who wish to replicate the study's results. More specifically, the researcher must systematically reflect upon why formative data were collected, how they were used, by whom they were used, and to what end. For example, to what extent did FE enable refinement to the implementation intervention such that the likelihood of encountering barriers in the future is adequately diminished? Or, in examining implementation issues across study sites, to what extent did FE provide information that led to modifications at individual sites? If the data and subsequent adjustments at individual sites were deemed critical to project success, upon broader dissemination to additional sites, what specific FE activities should be replicated, and by whom?

SUMMARY

Formative evaluation is a study approach that is often key to the success, interpretation, and replication of the results of implementation/QI projects. Formative evaluation can save time and frustration as data highlight factors that impede the ability of clinicians to implement best practices. It can also identify at an early stage whether desired outcomes are being achieved so that implementation strategies can be refined as needed; it can make the realities and black box nature of implementation more transparent to decision makers; and it can increase the likelihood of obtaining credible summative results about effectiveness and transferability of an implementation strategy. Formative evaluation helps to meet the many challenges to effective implementation and its scientific study, thereby facilitating integration of research findings into practice and improvement of patient care.

REFERENCES

1. **QUERI.** HSRD special solicitation for service directed projects (SDP) on implementation of research into practice (posted 2003). Available at: http://www.hsrd.research.va.gov/for_researchers/funding/solicitations/. Accessed August 10, 2005.
2. **Demakis JG, McQueen L, Kizer KW, Feussner JR.** Quality enhancement research initiative (QUERI): a collaboration between research and clinical practice. Med Care. 2000;38(suppl 1):I17–25.
3. **McQueen L, Mittman BS, Demakis JG.** Overview of the Veterans Health Administration (VHA) Quality Enhancement Research Initiative (QUERI). J Am Med Inform Assoc. 2004;11:339–43.
4. **QUERI.** QUERI program description (posted July 2003). Available at: http://www1.va.gov/hsrd/queri/. Accessed August 10, 2005.
5. **Cook TD, Campbell DT.** Quasi-Experimentation: Design & Analysis Issues for Field Settings. Boston: Houghton Mifflin Company; 1979.
6. **Campbell DT, Stanley JC.** Experimental and Quasi-Experimental Designs for Research. Boston: Houghton Mifflin Company; 1963.
7. **Isaac S, Michael W.** Handbook in Research and Evaluation: For Education and the Behavioral Sciences. San Diego: EdITS Publishers; 1982.
8. **Dzewaltowski D, Estabrooks P, Glasgow R, Klesges L.** Workgoup to evaluate and enhance the reach and dissemination of health promotion interventions (RE-AIM). Available at: http://www.re-aim.org/2003/whoweare.html. Accessed July 9, 2005.
9. **Glasgow RE, Lichtenstein E, Marcus A.** Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. Am J Pub Health. 2003; 93:1261–7.
10. **Stead M, Hastings G, Eadie D.** The challenge of evaluating complex interventions: a framework for evaluating media advocacy. Health Educ Res. 2002;17:351–64.
11. **Zapka J, Goins KV, Pbert L, Ockene JK.** Translating efficacy research to effectiveness studies in practice: lessons from research to promote smoking cessation in community health centers. Health Promot Pract. 2004;5:245–55.
12. **Ovretveit J, Gustafson D.** Using research to inform quality programmes. BMJ. 2003;326:759–61.
13. **Hulscher M, Laurant M, Grol R.** Process evaluation on quality improvement interventions. Qual Saf Health Care. 2002;12:40–6.
14. **Solberg LI.** Guideline implementation: what the literature doesn't tell us. Jt Comm J Qual Improv. 2000;26:525–37.
15. **Mittman B.** Creating the evidence base for quality improvement collaboratives. Ann Int Med. 2004;140:897–901.
16. **Dehar MA, Casswell S, Duignan P.** Formative and process evaluation of health promotion and disease prevention programs. Eval Rev. 1993;17:204–20.
17. **Rossi P, Freeman H.** Evaluation: A Systematic Approach. Newbury Park: Sage Publications; 1993.
18. **Bhola HS.** Evaluating "Literacy for Development" Projects, Programs and Campaigns: Evaluation Planning, Design and Implementation, and Utilization of Evaluation Results. Hamburg, Germany: UNESCO Institute for Education; DSE (German Foundation for International Development), xii; 1990.
19. **Patton MQ.** Evaluation of program implementation. Eval Stud Rev Annu. 1979;4:318–45.
20. **Altman DG.** A framework for evaluating community-based heart disease prevention programs. Soc Sci Med. 1986;22:479–87.
21. **Havas S, Anliker J, Damron D, Feldman R, Langenberg P.** Uses of process evaluation in the Maryland WIC 5-a-day promotion program. Health Educ Behav. 2000;27:254–63.
22. **Evans RI, Raines BE, Owen AE.** Formative evaluation in school-based health promotion investigations. Prev Med. 1989;18:229–34.
23. **Patton MQ.** Utilization-Focused Evaluation: The New Century Text. 3rd edn. Thousand Oaks, CA: Sage Publications; 1997.
24. **Patton MQ.** Qualitative Research & Evaluation Methods. 3rd edn. Newbury Park, CA: Sage Publications; 2001.
25. **Walshe K, Freeman T.** Effectiveness of quality improvement: learning from evaluations. Qual Saf Health Care. 2002;11:85–7.
26. **Wholey J, Hatry H, Newcomer K,** eds. Handbook of Practical Program Evaluation. San Francisco: Jossey-Bass Publishers; 1994.
27. **Forsetlund L, Talseth KO, Bradley P, Nordheim L, Bjorndal A.** Many a slip between cup and lip. Process evaluation of a program to promote and support evidence-based public health practice. Eval Rev. 2003;27:179–209.
28. Guide for Implementing Evidence-Based Clinical Practice and Conducting Implementation Research. Available at: http://www1.va.gov/hsrd/queri/implementation. Accessed August 10, 2005.
29. **van Bokhoven MA, Kok G, van der Weijden T.** Designing a quality improvement intervention: a systematic approach. Qual Saf Health Care. 2003;3:215–20.
30. **Rogers EM.** Diffusion of Innovations. 4th edn. New York: Free Press; 1995.
31. **Stetler C, Corrigan B, Sander-Buscemi K, Burns M.** Integration of evidence into practice and the change process: a fall prevention program as a model. Outcomes Manag Nurs Practice. 1999;3:102–11.
32. **Sales A.** Organizational readiness for evidence-based health care interventions. Available at: www.measurementsexperts.org/instrument;instrument_reviews.asp?detail=53. Accessed August 10, 2005.

33. **Luther SL, Nelson A, Powell-Cope G.** Provider attitudes and beliefs about clinical practice guidelines. SCI Nurs. 2004;21:206–12.

34. **Brown K, Gerhardt M.** Formative evaluation: an integrative practice model and case study. Personnel Psychol. 2002;55:951f.

35. **Legro M, Wallace C, Hatzakis M, Goldstein B.** Barriers to optimal use of computerized clinical reminders: the SCI QUERI experience. VA QUERI Quart. 2003;4:2.

36. **Wallace C, Hatzakis M, Legro M, Goldstein B.** Understanding a VA preventive care clinical reminder: lessons learned. SCI Nurs. 2004; 21:149–52.

37. **Stetler C.** The role of the organization in translating research into evidence based practice. Outcomes Manag Nurs Practice. 2003;7:97–103.

38. **Rycroft-Malone J, Kitson A, Harvey G, et al.** Ingredients for change: revisiting a conceptual framework. Qual Saf Health Care. 2002;11: 174–80.

39. **McCormack B, Kitson A, Harvey G, et al.** Getting evidence into practice: the meaning of 'context'. J Adv Nurs. 2002;38:94–104.

40. **Bradley E, Holmboe E, Mattera J, et al.** A qualitative study of increasing β-blocker use after myocardial infarction: why do some hospitals succeed? J Am Med Assoc. 2001;285:2604–11.

41. **Kimmel B.** Bridging the gap between knowledge and practice — the veterans administration pathway, Newsletter of the National Council of the University Research Administrators 2003–4;35.

42. **Dunlap M, Beyth R, Deswal A, Massie B, Saleh J, Kimmel B.** VA practice matters on treating chronic heart failure. VA Practice Matters. 2004;9:1–8. Available at: http://www.hsrd.research.va.gov/publications/internal/pm_v9_n1.pdf. Accessed September 9, 2005.

43. **Hawe P, Shiell A, Riley T.** Complex interventions: how "out of control" can a randomized controlled trial be? BMJ. 2004;328:1561–3.

44. **Kirchhoff K, Dille C.** Issues in intervention research: maintaining integrity. Appl Nurs Res. 1994;7:32–8.

45. **Santacroce SJ, Maccarelli LM, Grey M.** Intervention fidelity. Nurs Res. 2004;53:63–6.

46. **Boyd NR, Windsor RA.** A formative evaluation in maternal and child health practice: the partners for life nutrition education program for pregnant women. Matern Child Health J. 2003;7:137–43.

47. **Stetler C, Creer E, Effken J.** Evaluating a redesign program: challenges and opportunities. In Kelly K., ed. Series on Nursing Administration, Vol. 8. St. Louis: Mosby Year Book; 1996.

48. **Effken E, Stetler C.** Impact of organizational redesign. J Nurs Admin. 1997;27:23–32.

49. **Krumholz H, Herrin J.** Quality improvement: the need is there but so are the challenges. Am J Med. 2000;109:501–3.

50. **Willenbring ML, Hagedorn H.** Implementing evidence-based practices in opioid agonist therapy clinics. In: Roberts A., Yeager K., eds. Evidence-Based Practice Manual: Research and Outcome Measures in Health and Human Services 2004. New York: Oxford University Press; 2004: 340–7.

51. **Willenbring ML, Hagedorn H, Poster AC, Kenny M.** Variations in evidence-based clinical practices in nine United States Veterans Administration opioid agonist therapy clinics. Drug Alcohol Dependence. 2004;75:97–106.

52. **LaVela S, Legro M, Weaver F, Smith B.** Staff influenza vaccination: lessons learned. SCI Nurs. 2004;21:153–7.

53. **La Vela SL, Legro MW, Weaver FM, Goldstein B, Smith B.** Do Patient Intentions Predict Vaccination Behavior Over Time? Poster. Academy Health Annual Conference June 2004. San Diego, CA.

54. **Nazareth I, Freemantle N, Duggan C, Mason J, Haines A.** Evaluation of a complex intervention for changing professional behaviour: the evidence based out reach (EBOR) Trial. J Health Serv Res Policy. 2002;7:230–8.

55. **Kukafka R, Johnson SB, Linfante A, Allegrante JP.** Grounding a new information technology implementation framework in behavioral science: a systematic analysis of the literature on IT use. J Biomed Inform. 2003;36:218–27.

56. **Sanson-Fisher RW, Grimshaw JM, Eccles MP.** The science of changing providers' behaviour: the missing link in evidence-based practice. Med J Aust. 2004;180:205–6.

57. **Walker AE, Grimshaw J, Johnston M, Pitts N, Steen N, Eccles M.** PRIME—PRocess modelling in ImpleMEntation research: selecting a theoretical basis for interventions to change clinical practice. BMC Health Serv Res. 2003;3:22.

58. **Yin R.** Case Study Research: Design and Methods. Thousand Oaks: Sage; 1994.

59. **Patterson ES, Nguyen AD, Halloran JM, Asch SM.** Human factors barriers to the effective use of ten HIV clinical reminders. J Am Med Inform Assoc. 2004;11:50–9.

60. **Sharp ND, Pineros SL, Hsu C, Starks H, Sales AE.** A qualitative study to identify barriers and facilitators to implementation of pilot interventions in the Veterans Health Administration (VHA) Northwest Network. Worldviews Evidence-Based Nurs. 2004;1:129–39.

61. **Devers KJ, Sofaer S, Rundall TG** (Guest eds.). Qualitative methods in health services research: a special supplement to HSR. Health Services Res. 1999;34(part II):1083–263.

62. **Litchman J, Roumanis S, Radford M, et al.** Can practice guidelines be transported effectively to different settings? Results from a multi-center interventional study. J Comm J Qual Improv. 2001;27:42–53.