Evaluation of Public Health Demonstration Programs: The Effectiveness and Cost-Effectiveness of WISEWOMAN

ERIC A. FINKELSTEIN, Ph.D., M.H.A., 1 JOHN S. WITTENBORN, B.S., 1 and ROSANNE P. FARRIS, Ph.D., R.D. 2

ABSTRACT

Background: In today's healthcare environment, public health resources are scarce. Thus, interventions to improve the public's health must be rigorously evaluated to ensure that they make the best use of available resources.

Methods: The Centers for Disease Control and Prevention (CDC) provides a general framework for program evaluation. This paper presents additional details on several key evaluation areas within CDC's framework.

Results: Successful evaluations will be built into the program design; will be multifaceted, incorporating both quantitative and qualitative methods; will assess both process and outcome measures; and will engage stakeholders to ensure utility of results.

Conclusions: Well-planned evaluations can lead to less burdensome yet more effective assessment and better program performance and can increase the knowledge base for health promotion practice.

INTRODUCTION

The scarcity of public health resources in today's healthcare environment requires that interventions to improve the public's health be evaluated using rigorous scientific methods. Public health interventions that cannot demonstrate effective use of resources may not be implemented. Thus, evaluation designs must recognize and integrate the requirements of funding agents, ensure that intervention benefits can be accurately measured and conveyed, and ensure that areas for improvement can be continuously identified. Public health interventions are complex, and therefore proper evaluation is a challenge. Successful evaluations will be built into the program design; will be multifaceted, incorporating both quantitative and qualitative methods; will assess both process and outcome measures; and will engage stakeholders to ensure utility of results. Well-planned evaluations can lead to less burdensome yet more effective assessment and better program performance and can increase the knowledge base for health promotion practice.

The Centers for Disease Control and Prevention (CDC) provides a general framework for program evaluation.1 This paper presents additional detail on several key evaluation areas within CDC's framework. We emphasize how to apply the evaluation to demonstration programs, which require

2Centers for Disease Control and Prevention, Division of Nutrition and Physical Activity, Atlanta, Georgia.
Funding for this work was provided by the Centers for Disease Control and Prevention under contract number 200-97-0621. The views expressed in this paper are not necessarily those of the CDC or RTI International.
more detailed analyses than do fully funded pro-
grams and often necessitate collection of addi-
tional data. We focus primarily on analysis of the
effectiveness and cost-effectiveness of the CDC’s
Well-Integrated Screening and Evaluation of Wo-
men Across the Nation (WISEWOMAN) pro-
gram, an ongoing set of screening and lifestyle in-
tervention demonstration projects aimed at
reducing cardiovascular disease (CVD) risk fac-
tors among low-income, underinsured and unin-
sured women in 14 states and tribes. Specific ex-
amples are presented of how effectiveness and
cost-effectiveness analyses are being incorporated
into the WISEWOMAN demonstration program.
We first provide an overview of the WISE-
WOMAN comprehensive evaluation plan.

WISEWOMAN COMPREHENSIVE
EVALUATION PLAN

Because no single evaluation technique can ad-
dress evaluations of process measures as well as
intermediate and ultimate outcomes, different eval-
uation approaches should be coordinated to form
a comprehensive evaluation plan. Several evalua-
tion approaches make up the WISEWOMAN eval-
uation plan, which is concerned with whether the
program is successful at efficiently improving CVD
risk factors among participants and at meeting
other program goals and seeks to learn how and
why specific aspects of the program influence its
success. The WISEWOMAN evaluation plan doc-
ments the accomplishments of the program thus
far; supports ongoing quality assurance; ensures
implementation of interventions with fidelity to
design, measuring short-term, intermediate,
and long-term program outcomes; tests new ap-
proaches and innovations; and discerns best and
promising practices that can be used for program
improvement. The conceptual framework of the
WISEWOMAN program (Fig. 1) illustrates the pro-
gram components and opportunities for evaluation
questions. A draft evaluation matrix for WISE-
WOMAN is shown in Table 1. The matrix includes
structural and process determinants and short-
term, intermediate, and long-term outcomes across
individual, organizational, and community pro-
gram levels. The matrix makes clear the complex-
ity of the program and the multitude of evaluation
opportunities.

Limited resources require that evaluation of
WISEWOMAN take a phased approach. The fo-
cus of the early phase is on evaluating program
feasibility and acceptability and developing the
infrastructure required to evaluate program ef-
ectiveness and cost-effectiveness. The next two
sections discuss the effectiveness and cost-effec-
tiveness evaluation approaches of WISEWOMAN.
The focus of the second phase is on evaluating
best practices, which is addressed in another pa-
per in this Supplement.

Effectiveness evaluation

Although a program’s effectiveness cannot be
fully assessed until the participants have com-
pleted the intervention, effectiveness evaluation
must be designed before program implementa-
tion to ensure that the necessary preintervention
and postintervention data are collected. Thus, the
first task is to determine which data should be
collected. The data must be detailed enough to
answer key questions concerning intervention
benefits but not overly burdensome for the re-
spondents or those implementing the interven-
tion. For WISEWOMAN, participating local pro-
jects and the CDC have developed a standard set
of minimum data elements (MDEs) that must be
collected for all participants at baseline and at the
end of 1 year (Appendix). Projects submit the
MDEs to the CDC twice annually. Although most
projects also collect additional data and do their
own evaluation, the MDE data are consolidated
across programs, allowing for a more in-depth
analysis of the impact of WISEWOMAN than
would be possible with data from a single pro-
gram only.

The MDEs selected for WISEWOMAN focus on
specific risk factors (e.g., blood pressure, choles-
terol, and smoking rates). Many are identical to
those captured by the CDC’s Behavioral Risk Fac-
tor Surveillance System (BRFSS). This has sim-
plified the MDE pretesting and validation process
and allows for state-level comparisons of CVD
risk among WISEWOMAN participants and wo-
men in the general population.

The WISEWOMAN MDEs can be combined
into a summary measure that assesses overall
CVD risk: the change in 10-year probability of
coronary heart disease (CHD) over the 12-month
period following enrollment. Summary measures
are important because they allow evaluators to
assess the benefits of an intervention by using a
single measure that can be easily conveyed to pol-
cymakers. Ideally, a summary measure is broad
enough to allow for comparisons with similar interventions but narrow enough that an effective intervention will show statistically significant improvements given the available sample size. As a general rule, it becomes more difficult to accurately quantify a summary measure and a greater number of assumptions are required as the generalizability of the measure increases. The WISEWOMAN summary measure was chosen because it assesses overall reduction in CVD risk and is relatively easy to calculate on the basis of existing risk-scoring algorithms. \textsuperscript{5,6}

Summary measures assess changes in health status among those who receive an intervention. To determine if the changes are a direct result of the intervention, however, the participants' health status in the absence of the intervention must be determined. The standard approach is to compare the change in participants' health with that of a similar control group. For WISEWOMAN, however, this clinical trial approach is complicated by the lack of a true control group. In WISEWOMAN, ethical concerns mandate that all enrolled women, including those in control groups, receive the results of their baseline screening examination as well as the usual standard of care. Women who have dangerous CVD risk factor measures are referred to a physician and may receive education, medication, or additional follow-up care. Because women in the control group are provided with information and services, they may show different results than if they had received no services at all. As a result, without additional information or assumptions, WISEWOMAN evaluations can compare the effectiveness (and cost-effectiveness) of screening plus interventions with that of screening alone but cannot compare the effectiveness of the interventions with a baseline of no intervention.

In some instances, evaluators may be able to identify a suitable control group from outside the intervention. For example, WISEWOMAN projects have assessed general trends in CVD risk factors by using the CDC's BRFSS data. When incorporating data from outside sources, however, evaluators must consider the possibility that differences in sample selection and other potential biases may confound the comparisons. For example, because both the BRFSS and WISEWOMAN collect limited demographic data, we cannot be sure that the women in these two groups do not differ in unobservable characteristics that might affect their risk of CVD.
## Table 1. WISEWOMAN Evaluation Matrix

<table>
<thead>
<tr>
<th>Determinants of outcomes</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Program level</strong></td>
<td><strong>Structural</strong></td>
</tr>
<tr>
<td>Individual</td>
<td>Demographics</td>
</tr>
<tr>
<td>Risk factors</td>
<td></td>
</tr>
<tr>
<td>Family history</td>
<td>Risk factors</td>
</tr>
<tr>
<td>Family support and structure</td>
<td>Group or individual</td>
</tr>
<tr>
<td>Medical history</td>
<td>Referral</td>
</tr>
<tr>
<td>Efficacy</td>
<td>Follow-up</td>
</tr>
<tr>
<td>Organization</td>
<td>Policies</td>
</tr>
<tr>
<td>Programs</td>
<td></td>
</tr>
<tr>
<td>Clinical guidelines</td>
<td>Data management system</td>
</tr>
<tr>
<td>Efficacy</td>
<td>Coordination/referral</td>
</tr>
<tr>
<td>Community</td>
<td>Programs</td>
</tr>
<tr>
<td>Demographics</td>
<td>Number of community</td>
</tr>
<tr>
<td>Participation</td>
<td>organizations and other</td>
</tr>
<tr>
<td></td>
<td>government programs</td>
</tr>
<tr>
<td></td>
<td>and services offered</td>
</tr>
<tr>
<td></td>
<td>Outreach activities</td>
</tr>
</tbody>
</table>
An alternative approach to using supplemental data is to assume that in the absence of the intervention, risk factors remain unchanged. This approach requires interventions to show improvements in risk factors to be deemed successful. For example, suppose that women enrolled in WISEWOMAN showed no net change in 10-year CHD risk during their enrollment period. Given this result, one might assume that WISEWOMAN participation had no effect on CHD risk, but if this result occurred in the presence of an increase in CHD risk in the general population, then the intervention could be credited with avoiding these risks. However, if there had been a decrease in CHD risk in the general population, this could indicate that participation in the intervention actually increased CHD risk.

Experience has shown that it is often difficult to show statistically significant differences in a broad or comparable summary measure. For example, in an early analysis of the Massachusetts WISEWOMAN project, the difference in the summary measure was not statistically significant even though the screening and intervention group showed greater reductions in CHD risk than did the group that received screening only. An early analysis of the North Carolina WISEWOMAN project showed the same result. Thus, in addition to the summary effectiveness measure, evaluations should include supplemental indicators of effectiveness, such as changes in other outcome measures and behavior. In addition to changes in CHD risk, WISEWOMAN evaluates changes in blood pressure, cholesterol, smoking status, healthy eating indexes, and physical activity. Significant improvements in supplemental measures will provide additional evidence of the program’s benefits.

Cost-effectiveness evaluation

An effective intervention is a necessary but not sufficient condition for justifying the expansion of a demonstration program. An additional requirement is that a given intervention justify the use of scarce public health resources. Cost-effectiveness analysis, defined as the costs required to yield a specific, comparable outcome, addresses this issue and can play an important role in determining funding priorities.

When constructing a cost-effectiveness ratio, it is important to consider the question the ratio is intended to answer. For example, the summary measure of effectiveness for WISEWOMAN is the 10-year probability of CHD. For WISEWOMAN projects that enroll a comparison group, the cost-effectiveness analysis addresses the incremental cost-effectiveness of screening plus intervention compared with screening alone and can be calculated as follows:

$$\text{Cost-effectiveness ratio} = \frac{\text{COST(IG)} - \text{COST(CG)}}{\Delta \text{CHD_RISK(IG)} - \Delta \text{CHD_RISK(CG)}}$$

where CG represents the control group (screening only) and IG represents the intervention group (screening plus intervention). A cost-effectiveness analysis of the early version of the Massachusetts WISEWOMAN project found that it cost an additional $637 to achieve a (statistically insignificant) 1-percentage point larger decrease in the 10-year probability of CHD in the screening plus intervention group than in the screening only group. Incremental change for either group (screening or screening plus intervention) can be calculated by assuming that CHD risk is flat for nonparticipants or by estimating changes in CHD risk for nonparticipants by using BRFSS or similar data. Because the cost of providing no screening or intervention is zero, a cost-effectiveness ratio that compares screening, or screening plus intervention, with no intervention can be easily calculated by using the cost-effectiveness formula. This ratio can then be compared with the cost-effectiveness ratios of similar interventions to determine a program’s relative cost-effectiveness.

Although the formula for determining cost-effectiveness is simple, deciding which costs to include and accurately quantifying them often is difficult. Issues related to effectiveness focus primarily on identifying an appropriate measure of effectiveness. The primary issue concerning costs centers on which costs to include and how to quantify them. In general, cost-effectiveness analyses should include opportunity costs of participation, defined as the opportunities foregone as a result of participating in intervention-related activities; direct dollar outlays associated with the implementation; and the dollar value of program resources not reflected in program budgets, such as volunteer labor and on-hand or donated facilities and equipment. The total of these costs, termed societal costs, is required for proper cost-effectiveness analysis.
Although direct dollar outlays are readily available from program budgets and the valuation of free and on-hand resources can be easily estimated, measurement of participant opportunity costs is more complicated. As is the case with many interventions, participation does not end when a woman leaves an intervention session. For WISEWOMAN, it would be extremely expensive to quantify all the additional time that women spend performing related activities, such as reading food labels, cooking healthier meals, and walking. For demonstration projects, it may be appropriate to ignore opportunity costs in the demonstration phase and to simply explore how many additional resources (regardless of payer) would be required to fund an expansion of the existing intervention. Only when an intervention proves to be effective should additional resources be dedicated to quantifying costs from a broader societal perspective.

Quantifying incremental costs, which are the relevant costs included in cost-effectiveness ratios, is not straightforward. Some evaluators make the simple assumption that if the size of the program were doubled, a doubling of the original budget would be required to fund the expansion. However, this is unlikely to be the case. The cost of expanding a successful demonstration project is generally less than the cost of providing the original demonstration because many activities do not need to be repeated. Economists define these one-time costs as sunk costs. For example, WISEWOMAN demonstration projects spend substantial resources developing nutrition and physical activity interventions, designing forms, developing training programs, and conducting other activities that would not need to be repeated if the intervention were expanded. These one-time costs should not be included in the numerator of a cost-effectiveness ratio that is being used to consider whether to expand a demonstration project. On the other hand, most successful interventions rely on existing infrastructure and collaboration with other agencies to deliver some services. For example, some WISEWOMAN projects refer participants to existing smoking cessation classes. Although the cost may not be funded out of the WISEWOMAN budget, the classes represent real expenditures that should be included in the cost-effectiveness analysis.

In many cases, it is not obvious whether certain costs should be included. If a woman enrolls in WISEWOMAN and is subsequently found to have heart disease requiring major surgery, it is unclear if the surgery and follow-up care costs should be included in the societal cost of the demonstration project. One solution is to exclude from the analysis any costs that are not expected to recur. Stated another way, if one individual has a high-cost random experience (e.g., a medical intervention), the costs can be excluded, but if many individuals have the same experience, the occurrence ceases to be random and the costs need to be considered.

Once evaluators have identified appropriate project costs, they must determine how to measure them. A straightforward approach is to measure costs at the primary activity level. The activity-based approach assesses how much it costs to provide each primary activity and identifies the components of these costs (including sunk costs). If implemented correctly, this approach can capture most project-related costs and facilitate the cost-effectiveness analysis. For example, WISEWOMAN has three primary activities—outreach and follow-up, screening, and intervention—and a catchall activity called administration. Outreach and follow-up consist of activities associated with enrolling women and promoting retention. Screening includes activities associated with collecting medical information and communicating assessment results. Intervention includes activities associated with providing the nutrition and physical activity interventions. Administration includes all remaining activities that cannot be allocated to one of the three primary activities. For the Massachusetts evaluation noted previously, the costs associated with all three activities averaged $603 per participant.

Within each activity category, there are three types of costs (that may or may not be sunk costs), (1) (noncontract) labor costs, (2) contracted service costs; and (3) costs associated with facilities, materials, and supplies. Noncontract labor costs consist of the dollar value (salary plus benefits) of the time that employees spend performing demonstration activities. These costs should be measured (e.g., via time sheets) for all individuals whose labor is not captured through another mechanism (i.e., billed directly to the project). Contracted service costs include costs for demonstration activities provided by outside entities. For example, in most states, physicians who perform WISEWOMAN screenings submit a bill for services rendered. Documenting contracted service costs is straightforward but requires sites to keep track of the bills associated with contracted services and to map the bills to the appropriate activity. Finally,
the costs of facilities, materials, and supplies include costs for promotional materials and materials or supplies purchased to support intervention activities (e.g., cookbooks and pedometers).

DISCUSSION

Effectiveness and cost-effectiveness analyses are two important aspects of a comprehensive evaluation. It is these outcome measures that may be the most persuasive to stakeholders and policymakers. However, these measures do not identify intervention strengths and weaknesses, and it is often difficult to show statistical significance in the summary outcome measures used in cost-effectiveness ratios, even for interventions that are truly effective. Also, cost-effectiveness ratios are not the only measures that policymakers consider. Thus, cost-effectiveness analysis should be only one component of a comprehensive evaluation. Before summary outcomes can be generated, an evaluation must assess the intervention design and implementation processes to ensure program success. Before outcome measures can be fully understood and interpreted, an evaluation must assess intermediate outcomes and other effects of the evaluation. The complexity and context-dependent nature of health promotion programs necessitate that cost-effectiveness analysis be implemented in conjunction with other evaluation techniques, including process evaluation and additional outcome evaluation.

In this paper, we have described the effectiveness and cost-effectiveness evaluation methods for WISEWOMAN because answering questions related to effectiveness and cost-effectiveness is a priority in an era in which public health interventions must be able to convincingly prove their worth to maintain funding. However, WISEWOMAN incorporates many comprehensive evaluation elements in addition to the effectiveness and cost-effectiveness analyses. These include process, structural, and supplemental outcome measures used to assess design, implementation, and effects at individual, organizational, and community program levels.

Process evaluation focuses on the operation of a program—what is working and what is not—and has been defined as relating to assurance of quality of practice or as an assessment of how a program is implemented. Process evaluation can assess how a program is structured and implemented and what activities are provided, under what conditions, by whom, and for whom. Critical process indicators collected by WISEWOMAN include how long it takes to fully field the intervention, enrollment rates, participant dropout rates, staff turnover, and allocation of budgetary funds. Process indicators can also be qualitative, intended to provide a detailed understanding of both the internal dynamics of a program and its interaction and relation with the environment. For example, WISEWOMAN projects are required to describe instances of community engagement, such as the use of existing community engagement, such as the use of existing community engagement, such as the use of community resources in the intervention, both for assessing intervention design and for ensuring implementation quality and efficiency. Process measures may alert project administrators and evaluators to problems in implementation, problems in recruitment, or intervention design flaws that could jeopardize the ultimate success of an intervention.

The WISEWOMAN evaluation plan also incorporates additional outcome measures. These short-term, intermediate, or long-term outcome measures evaluate intervention effects not captured by the summary outcome measure. These measures can provide earlier or more descriptive evidence of intervention effects than is captured by the summary outcome measure. Impact evaluation can provide important individual-level data, such as tracking participants’ precursors to change. Some WISEWOMAN projects measure changes in knowledge, attitudes, and beliefs (KAB) and in self-efficacy, which are expected to precede any lifestyle or behavior changes caused by the intervention. Other projects include participants’ stage of change, in which the participants are classified within a process of change delineated into a series of sequential stages. Not only can stages of change be an effective tool for tailoring interventions to each individual’s readiness for change, but they also can detect early evidence of change or identify barriers to change. Impact measures included in the WISEWOMAN MDEs include changes in nutrition, physical activity, and smoking behavior. In addition to individual-level impacts, the evaluation will also attempt to document impacts on organizations and communities because these can constitute a major, lasting achievement of health intervention projects. Thus, a goal of WISEWOMAN is to encourage improvement in efficiency and effectiveness among the sponsoring agencies and community partners and to improve provider norms. Community impacts can constitute effects of even greater scope than organizational change.
stages of the WISEWOMAN evaluation will assess changes in overall population health and impacts on community health delivery systems, resources, and the environment using such measures as community-level indicators.15,16

The evaluation techniques briefly described here identify those aspects of a program that are more or less successful at effecting behavior change, and why. In the context of the early phases of WISEWOMAN program evaluation, they serve to bolster the ultimate products of the evaluation: the effectiveness and cost-effectiveness analyses. Process evaluation can help ensure that an intervention operates as it should to produce the desired outcomes by facilitating improvement and providing an improved understanding of how the intervention functions. Impact evaluation is used to document the early effects of an intervention as well as outcomes that may not be captured in the summary outcome measure selected for the cost-effectiveness analysis. It is the cost-effectiveness outcome evaluation, however, that will likely be the ultimate measure of a program’s worth. The outcome evaluation must correctly capture the most important effect of the intervention in a manner that is accurate while still being comparable to the outcomes of other health interventions. The outcome evaluation must also accurately measure the resources required to achieve this outcome. A truly comprehensive evaluation will provide accurate, unbiased information to ensure that an intervention operates to its full potential and to appraise and communicate the intervention’s eventual level of success. Ultimately, the evaluation will serve as the basis for determining whether to expand a demonstration project and how to tailor the intervention to improve health among participants and other affected parties.

REFERENCES


16. Department of Health Services, School of Public Health, University of Washington. Community-level indicators [Internet]. Available from faculty.washington.edu/cheadle/cli/
APPENDIX. WISEWOMAN MINIMUM DATA ELEMENTS (MDEs)

Minimum data elements
- Date of birth
- Ethnicity and race
- Education (highest grade completed)
- Height
- Weight
- Blood pressure information
- Cholesterol (HDL and total, fasting status)

Questions
- Have you ever been told by a doctor, nurse or other health professional that your blood cholesterol is high?
  - that you have high blood pressure?
  - that you have diabetes?
- Are you currently taking medication for high cholesterol?
  - for high blood pressure?
  - for diabetes?
- Do you now smoke cigarettes?