PROGRAM EVALUATION IN THE ABSENCE OF GOALS: A COMPREHENSIVE APPROACH TO THE EVALUATION OF A POPULATION-BASED BREAST CANCER SCREENING PROGRAM

Shawna L. Mercer
Vivek Goel
Institute for Clinical Evaluative Sciences in Ontario
North York, Ontario
and
Department of Preventive Medicine and Biostatistics
University of Toronto
Toronto, Ontario

Abstract: The aim of the province-wide Ontario Breast Screening Program (OBSP) is to reduce breast cancer mortality through providing screening for breast cancer to all Ontario women 50 and over. The authors identify the components of the OBSP and the organizational, political, and financial challenges and biases affecting its evaluation. They then argue that because key players have not yet established program goals from which process and outcome indicators can be identified, a goal-directed evaluation is not currently possible. They therefore develop a framework to permit the identification, categorization, and prioritization of all factors that need to be addressed in a comprehensive evaluation of the OBSP. This evaluation approach may be of assistance to evaluators of other complex programs about which consensus on program goals has not yet been attained.

Résumé: Le Programme ontarien de dépistage du cancer du sein (PODCS) vise à réduire la mortalité découlant d’un cancer du sein en assurant des examens à toutes les femmes de l’Ontario âgées de 50 ans et plus. Cet article indique les composantes du PODCS ainsi que les biais et les défis organisationnels, politiques et financiers pouvant influer sur son évaluation. Il précise par la suite qu’il est actuellement impossible de procéder à une évaluation des objectifs atteints étant donné que les responsables n’ont pas encore déterminé les objectifs du programme pouvant permettre la mise au point d’indicateurs. Un cadre de travail a donc été élaboré en vue de permettre la détermination, la catégorisation et la mise en ordre...
Breast cancer is the most common cancer occurring in Canadian women as well as the leading cause of cancer deaths in females (National Cancer Institute of Canada, 1991). Each year, approximately 14,400 Canadian women contract breast cancer and 5,100 die from the disease (National Cancer Institute of Canada, 1991). Research has shown that screening of asymptomatic women in order to detect breast cancer at the earliest possible stage is effective in reducing breast cancer mortality in women 50 and above (Day & Miller, 1988; Eddy, 1989; Rutqvist et al., 1990). A Canadian expert working group, convened by the Conference of Federal–Provincial Deputy Ministers of Health and consisting of key volunteer and professional organizations and government, recommended that Canadian provinces consider as high priority the development and implementation of programs providing breast screening every two years to women 50 and over (Workshop Group, 1989).

In accordance with these recommendations, the Ontario Breast Screening Program (OBSP) has been initiated as a service of the Ontario Cancer Treatment and Research Foundation (OCTRF) funded by the Ontario Ministry of Health. When fully implemented, the OBSP plans to provide screening for all Ontario women 50 and over through approximately 36 screening centers located in easily accessible settings across the province.

This article describes the major components of the OBSP, problematic issues that affect its evaluation and preclude the utilization of a goal-directed evaluation approach, and the resulting framework that has therefore been developed in order to design a comprehensive evaluation of the program.

PROGRAM DESCRIPTION

There are six major components to the program. Figure 1 illustrates their interrelationships.
Figure 1

Relationships between Components of the Ontario Breast Screening Program

- Health Promotion
- Professional Education
- Recruitment/Retention
- Screening
- Follow-up
- Evaluation/Research Partnerships
Health Promotion

The program’s health promotion component aims to increase awareness and knowledge of breast health and the importance of breast screening, as well as to promote attitudes and behaviors that will result in women’s greater participation in the program. Community mobilization techniques (Stunkard, Felix, & Cohen, 1985) are utilized to ensure that screening centers meet the needs of the communities they serve and that the communities have a sense of ownership of their local program.

Professional Education

Professionals involved in the operation of the program take part in educational and quality assurance activities to ensure efficient program delivery. Professional groups external to the program but involved with the referral and follow-up of women are offered written and verbal information about the OBSP and the opportunity for continuing education in areas related to the program.

Recruitment

A multifaceted approach is being used to recruit women into the OBSP. Women may be referred either by their physician or through self-referral. The Health Promotion and Professional Education program components are thus crucial to successful recruitment. Recruitment plans also include issuing letters of invitation to women to attend for screening.

Screening

The core of the program is the screening itself. This consists of two modalities: physical examination of the breasts by a trained nurse examiner, and two-view mammography with dedicated equipment. A quality assurance system seeks to ensure that the screening interventions are applied consistently with good quality. Breast self-examination videos and pamphlets are also available at the centers.

Follow-up

The screening process will classify women as normal or as requiring further assessment. An assessment center has been established in
one screening center on a pilot basis. In other screening centers either women needing follow-up will be referred back to their family doctors, or the screening center staff may arrange follow-up appointments for women at the request of their family physician. Follow-up may include clinical examination, diagnostic mammography, ultrasound, surgical consultation, fine-needle aspiration, biopsy, or open biopsy.

Evaluation and Research Partnerships

Organized breast screening is a new service intervention in Canada, and its delivery in Ontario involves many innovative approaches. Thus, evaluation and research are key components in determining the utility of the OBSP. Because the OBSP is a service program rather than a research project, evaluation will be ongoing and carried out largely by program staff, whereas research will involve program staff collaborating with researchers external to the program.

PROBLEMATIC AREAS AFFECTING OBSP EVALUATION

A number of characteristics of the program make its evaluation complex. They can be categorized as follows.

Organizational Context of the OCTRF

Breast screening represents a new thrust for the OCTRF, which has traditionally been treatment oriented. The OBSP is a secondary prevention program providing services to healthy women who first need to be convinced of the benefits to be gained through screening. This engenders a major shift for an organization that has historically provided specialty care to cancer patients.

In addition, the OCTRF operates with a decentralized management system that has been applied to the Breast Screening Program. Evaluation of a provincial program typically requires further standardization of policies and procedures than has occurred to date within the OBSP.

Program-Specific Factors

The evaluation is also complicated by the fact that the OBSP is multifaceted. Its successful implementation will require close coopera-
tion between individuals from many different backgrounds. In particular, applying a secondary prevention maneuver that relies on the medical model of health care delivery within the context of a health promotion intervention is unique. The evaluation has to be tailored to examine issues in each of these divergent areas.

Additionally, the program is engaged in building coalitions that involve many partner agencies in supporting the program and ensuring its sensitivity to local needs. The evaluation must consider the program-related activities of each of these groups.

External Factors

The OBSP received its funding approval and has undergone its entire implementation to date in the midst of the current recession. As a result, the program’s global budget has been severely restricted, necessitating careful prioritization of all evaluation requirements.

Ontario has seen a rapid increase in unorganized screening mammography in the last 10 years—much of it in women for whom screening mammography has not been recommended, or at intervals shorter than those recommended by the Canadian expert working group. Quality assurance procedures such as those in place within the OBSP may not be applied in unorganized settings. For all of these reasons, mammography outside the program may not be judiciously executed. However, because of their effects on morbidity, mortality, and OBSP recruitment, the outcomes of unorganized screening also have to be evaluated.

A further complicating issue the evaluation faces is its dependency on information that is beyond the scope of the program. The program itself ends when the woman is found through screening to be normal or when she is referred for follow-up. Information on the outcomes of the screening—procedures performed on women referred for assessment, any delays in the follow-up process, and the diagnosis of cancer—will have to be collected from various external sources.

Factors Affecting the Evaluation of Goal Attainment

The traditional evaluation approach uses a program’s goals and objectives to identify process and outcome indicators for which data are then collected and analyzed in order to judge program success and delivery (Attkisson & Hargreaves, 1979; Borus, Buntz, & Tash, 1982;
Shortell & Richardson, 1978). To date, this methodology has not been possible within the OBSP.

The OBSP was originally designed by taking the findings of a number of epidemiological studies with experimental and control populations and extrapolating these results to fit the entire Ontario population. This permitted identification of the number of women that would need to undergo regular screening in order to achieve a particular reduction in mortality. It was later recognized that ensuring sufficient participation of the target population in the program would be difficult. The health promotion program component was added to address this need. The original program proposal, however, provided only a brief and sketchy discussion of the role of health promotion.

As a result of the above, program providers have varying conceptions of the OBSP. Although there is consensus that the overall aim of the program is to reduce mortality from breast cancer, the program goals and objectives, along with their relative priorities, have not yet been finalized. Work is ongoing to formalize the goals and objectives; however, the program evaluation approach has had to be developed concurrently in order that data recognized as essential not be lost.

Despite the fact that it is possible to measure mortality from breast cancer in Ontario using statistics collected by the Registrar-General’s Office, it will be difficult to show what changes are attributable to breast screening, for several reasons. First, it may take a number of years to detect a significant difference in mortality between women who receive screening and those who do not (Shapiro, Venet, Strax, Venet, & Roeser, 1982). Moreover, this lag period is counted from when the program achieves full or close to full coverage of the population, which in Ontario may not be reached for several years. Second, it will be difficult to identify an adequate comparison group that does not receive screening. Most jurisdictions are already heavily contaminated by community mammography, and the current increases in breast cancer incidence and mortality (Clarke, Goel, & Fleury, 1991) may mask changes that occur as a result of screening and reduce the appropriateness of a historical cohort. Additionally, using unscreened women in Ontario as controls may introduce considerable selection bias, as women who attend screening voluntarily have different health behaviors from those who do not (Vernon, Laville, & Jackson, 1990).

Two other types of biases may occur in examining follow-up data from screening programs (Day, Williams, & Khaw, 1989). Lead-time bias
can result because of earlier detection of a breast lesion without any change in prognosis. Survival time thus appears to be longer, when in fact these patients die at the same time they would have without early detection and are only made to live a longer period of time with the knowledge of their disease. Length bias can occur when tumors with longer preclinical phases, such as those that are slower growing or less aggressive, are preferentially detected in a screening program. Because these tumors have a better prognosis, the screened cases seem to have better survival.

OVERALL APPROACH TO EVALUATION

As a result of these problematic issues, the traditional evaluation approach has had to be modified in order to meet the immediate and long-term evaluation needs of the OBSP. Most notably, the evaluation cannot at the present time be goal driven. A number of outcome indicators that must be included in evaluating any breast screening program’s progress toward mortality reduction have, however, already been identified in the literature (Day, Williams, & Khaw, 1989; Gray, 1990; Paci et al., 1990). Moreover, despite the fact that comparable information regarding process indicators does not exist, examination of OBSP structure, implementation, and operations has revealed many factors important to the evaluation of process, as well as additional outcome measures. A framework was sought that would coordinate all of this information in a logical and useful fashion.

Development of a Framework for the Evaluation

The development of a logic model to illustrate program components, what they are expected to produce, and how they relate to each other has been identified as a crucial step in the evaluability assessment of any program (Rutman & Mowbray, 1983). Formalization of the OBSP evaluation framework thus began with the preparation of an OBSP logic model (Figure 2). This entailed taking each of the identified program components and listing both the service elements involved in carrying them out and their expected outputs. The immediate to long-term goals and effects anticipated as a result of each component were then identified, along with their interrelationships. The majority of OBSP stakeholders agree about the desired effects of the program and its various components. Thus, entitling the final category “goals and effects” circumvents the fact that final consensus on the goals of the OBSP and their prioritization has not yet been achieved.
Figure 2
A Logic Model for the Ontario Breast Screening Program
The OBSP logic model appears to be quite complex. This is partly because of lack of formalization of program goals, but is also due to the fact that the program is a multifaceted integration of the medical model and health promotion and, as such, has many desirable and interdependent endpoints.

The next step in establishing an evaluation framework was to develop a grid providing a comprehensive listing of the measures to be evaluated for all aspects of the OBSP (examples for each program component are provided in Figure 3). The grid considers each program component separately and identifies the service elements and associated outputs that are involved in carrying out that component. Process and outcome indicators are then listed for each service element. Lastly, the grid identifies how data can be collected for each indicator.

The logical progression and layout of the program logic model and the evaluation grid can be seen in Figures 2 and 3. Taking the health promotion component as an example, its service elements include social marketing and education. Outputs for the health promotion component include awareness and education materials and initiatives. Process indicators for these service elements include identification of which materials are the most and the least utilized and effective among various populations. Data on these indicators can be collected through the use of a special study (which might include focus testing and observational investigations). Outcome indicators include the level of awareness of the importance of early detection among various populations. Awareness can be measured through surveys performed at various intervals.

In order to facilitate coordination and prioritization among the large number of evaluation projects and studies identified through the grid, the indicators within each specific program component were assembled in tabular format by data collection strategy. For each component, this yielded lists of the following: (a) data to be collected through ongoing monitoring, (b) any indicators requiring the use of additional data sources, (c) data to be gathered through the use of special studies, and (d) indicators to be measured through surveys. Within the survey and special study categories, indicators that could be addressed through the same project were then grouped together. These groupings of indicators provided themes and parameters for specific surveys and special studies.
### An Evaluation Grid for the Ontario Breast Screening Program

<table>
<thead>
<tr>
<th>Component</th>
<th>Service Elements</th>
<th>Intermediate Term/ Process Indicators</th>
<th>Data Collection</th>
<th>Longer Term/ Outcome Indicators</th>
<th>Data Collection</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Health Promotion</strong></td>
<td>Social Marketing/Education</td>
<td>- identification of which mediums or specific program materials are: - most and least utilized - most and least effective</td>
<td>Special Study</td>
<td>- Awareness of importance of early detection among: - Target age group - population subgroups - General public</td>
<td>Monitoring &amp; Special Study</td>
</tr>
<tr>
<td></td>
<td>Pamphlets, Brochures</td>
<td></td>
<td></td>
<td>- Support for OBSP: - financial - equipment, supplies, etc. - staffing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Talk Shows</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Advertisements:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Newspaper, TV</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Community Mobilization</td>
<td>- No. of coalitions established - Membership of coalitions - Activities of coalitions - Tangible benefits of coalitions</td>
<td>Monitoring</td>
<td>- Awareness of the importance of early detection among health professionals - Referral of women to the program by their doctors</td>
<td>Monitoring &amp; Special Study</td>
</tr>
<tr>
<td></td>
<td>Coalition Building (Community Committees)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Provincial, Regional</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Professional Education</strong></td>
<td>Teaching/Sponsoring/ Coordinating</td>
<td>- Factors deterring doctors from referring patients to OBSP - Types of doctors referring women to the program - General practitioners, others</td>
<td>Survey &amp; Special Study</td>
<td>- Coverage within OBSP = total eligible population - Demographics of women screened vs. eligible</td>
<td>Survey &amp; Census Data</td>
</tr>
<tr>
<td></td>
<td>Educational Courses</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Forums</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Conferences</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Recruitment/ Retention in Screening Program</strong></td>
<td>Generation of Letters of Invitation</td>
<td>- Compliance with letter of invitation - Effect of different invitation letter types on recruitment - Barriers to screening</td>
<td>Monitoring</td>
<td>- No women screened by OBSP</td>
<td>Monitoring &amp; Census Data</td>
</tr>
<tr>
<td></td>
<td>Personalized</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>General</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Demographic Questionnaire</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Screening</strong></td>
<td>Physical Examination</td>
<td>- Satisfaction of women with screening center visit: - Center-specific factors - Intentions to return, to tell friends to attend</td>
<td>Client Satisfaction</td>
<td>- No. of women referred for assessment: - by physical exam alone - by mammography alone - by both modalities</td>
<td>Monitoring &amp; Census Data</td>
</tr>
<tr>
<td></td>
<td>- Performed by Trained Nurse</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Two-View Mammography</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Films Read by a Radiologist</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Follow-up of Screened Women</strong></td>
<td>Forms Requesting Follow-up Information</td>
<td>- Delays to each step in the follow-up process - No. &amp; types of procedures performed during follow-up - Differences in specialists' recommendations</td>
<td>Monitoring &amp; Special Study</td>
<td>- No. of breast cancers detected</td>
<td>Monitoring, OCR Data</td>
</tr>
<tr>
<td></td>
<td>Liaison with Ontario Cancer Registry (OCR)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Breast Cancer Incidence Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Breast Cancer Mortality Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Prioritization of Possible Evaluation Projects

The final tabular listing of program indicators according to data collection methodology is a straightforward document from which program staff, external stakeholders, and consultants are able to identify the scope of the data that would need to be collected for a comprehensive evaluation of the OBSP. It further enables these same people to prioritize all monitoring data and possible projects according to overall importance and the most appropriate timing for data collection. During prioritization, the program logic model and evaluation grid are made available to illustrate how all of the data tie together and the rationale behind measuring each particular indicator.

Prioritization of evaluation studies and monitoring data for the OBSP has been based on a number of criteria that include the overall program budget and the specific budget allotted to evaluation, the implementation schedule for the program, staffing for program delivery and evaluation, the impact of particular evaluation strategies on screening center operations, current and planned levels of computerization, and the concerns and requirements of external stakeholders and funders.

Other criteria for prioritization have evolved from the problematic issues identified earlier. All components of this multifaceted program need to be evaluated. Evaluation must take into account the regional tailoring of some program components to fit local needs. What and how much data can be effectively collected from external data sources must be determined. Lastly, and of paramount importance, is the need right from the start to collect data on intermediate indicators—those indicators that will reveal whether the program is on its way to producing the significant mortality reduction that will not be evident for at least seven to ten years.

IMMEDIATE EVALUATION NEEDS

The prioritization process suggested the following to be the immediate evaluation needs for the OBSP.

Ongoing Program Monitoring and the Use of Additional Data Sources

Ongoing monitoring will remain an essential element of data collection for each program component. Health promotion record-keeping includes a description of all activities undertaken, along with a com-
parison of potential and actual numbers of participants or individuals reached. The utilization of awareness and education materials by health professionals and the general public is also monitored.

Monitoring data collection further includes the numbers of women who attend for screening, and their age category. Coverage—the percentage of women in the target population who complete screening—is one of the single most important determinants of the outcome of a screening program (Day, Williams, & Khaw, 1989; Gray, 1990) and will be calculated by comparing the number of attendees to estimates of the total number of eligible women in the population. Such estimates will be garnered largely from Statistics Canada census data. The age data will permit standardization of the expected rates of breast cancer incidence and mortality.

The demographic characteristics of women in different geographic regions will also be procured from census data. These characteristics will be compared to detailed demographic information collected on women that attend screening in order to determine profiles of attenders and nonattenders.

Adherence to quality assurance standards is monitored for each of physical examination, mammography (technologists and equipment), and the interpretation of mammograms. Women’s satisfaction with their screening visit is also examined regularly.

Information and reports on procedures completed, diagnoses, and elapsed time for all women referred for follow-up is being requested from assessment centers and family physicians, and also from hospitals and specialists when necessary. Operative and pathological reports permit an analysis of tumor size and the stage distribution of cancers detected by screening. Comparison of the observed distributions of stage and tumor size to the expected provide two of the most important predictors of long-term mortality reduction (Day, Williams, & Khaw, 1989).

A working relationship is being established with the Ontario Cancer Registry in order to ascertain breast cancer incidence and mortality in women that attend screening (including interval cancers) as well as in those that do not. Careful attention will also be given to sensitivity (the proportion of women judged to be abnormal on screening who actually have breast cancer) and specificity (the proportion of women whose screening is normal who truly do not have cancer) and thus also to false negatives and false positives.
Lastly, monitoring will include collecting data on direct program costs through the cost accounting system of the OCTRF. Along with future studies of direct and indirect costs, this will enable the program to determine if it is using provincial health resources efficiently.

Special Studies and Surveys

A survey of the awareness, knowledge, attitudes, intentions, and behaviors toward breast screening of women in the target age group is currently underway. A survey of the knowledge, attitudes, and referral and breast cancer control practices of physicians is also being undertaken.

One special study to be initiated in the near future will compare the various recruitment strategies. Another will examine whether women with abnormalities receive appropriate and expedient follow-up and will compare the pilot assessment center site to the other OBSP regions. A further special study will look at the importance of nurse examiners to the program—not only in detecting abnormalities, but also evaluating whether they provide significant support, reassurance, and education to women undergoing screening. Another study being considered will seek to determine the levels of anxiety and quality of life in women who have attended for screening.

CONCLUSIONS

This article provides an overview of the challenges presented by the evaluation of a program as large and complex as the OBSP. It explains why a goal-directed evaluation is not possible at this stage of the OBSP’s development and describes the development of a framework to permit identification and categorization of all factors to be addressed in a comprehensive evaluation of the OBSP. The article then outlines how these factors have been prioritized in order to delineate the OBSP’s immediate evaluation needs.

Aside from its relevance to the planning, administration, and evaluation of screening programs, the present article describes an approach to evaluation that may prove helpful to evaluators of widely differing programs. The OBSP is illustrative of many current programs that are multifaceted in nature and that, owing to political and other pressures, are often implemented prior to the completion of program planning or before stakeholders reach consensus on all goals and objectives and their relative priorities. As is the case with the
OBSP, it may be impossible to postpone data collection in these programs until goals have been finalized if program effectiveness is to be evaluated in a timely manner.

Evaluation of such programs may be facilitated with the aid of the framework provided in the current article. As a first step, evaluators must identify the components of the specific program to be evaluated. A program logic model should subsequently be prepared. A grid then needs to be developed that identifies all indicators that could potentially be evaluated for each program component and indicates how data could be collected for each indicator. The information garnered through this grid should next be assembled for each component in tabular format by data collection strategy, and indicators that could be addressed in the same study should be grouped together. Evaluation studies and ongoing monitoring data identified in this tabular listing can then be prioritized by taking into account organizational, program-specific, budgetary, and external factors. The result of this approach to evaluation is purposeful and directed data collection that furnishes answers to those questions that are most critical to the short- and long-term evaluation of program delivery and success.

REFERENCES


