Use of an interrupted time-series design to evaluate a cancer screening program

Robert Michielutte, Brent Shelton, Electra D. Paskett, Cathy M. Tatum and Ramon Velez

Abstract

An alternative approach to intervention–control designs to evaluate community health education studies is to use a quasi-experimental design in which the outcomes of interest are examined over time in the intervention unit. The Forsyth County Cancer Screening Project (FoCaS) was a comprehensive clinic- and community-based education program to increase screening for cervical cancer and breast cancer among low-income women. This paper reports the use of piecewise regression accounting for potential effects of auto-correlation in the data to evaluate the effectiveness of the project in increasing mammography screening. Data for the evaluation of trends in screening consisted of all mammograms performed during the period of May 1992 through June 1995 at the Reynolds Health Center in Forsyth County, North Carolina. The results suggested that the FoCaS project was effective in increasing mammography screening among women age 40 or older in the study population. Analysis of the trends by age indicated that the program had differential effects on women age 40–49 and 50 or older. The results demonstrate that analyses of the type presented here can either complement or serve as an alternative to more traditional intervention–control analyses.

Introduction

Prevention and early detection strategies play an important role in reducing incidence and mortality from cancer in the US (Public Health Service, 1991). Public health intervention programs can be effective in modifying lifestyle behaviors, reducing exposure to carcinogenic environmental factors, and increasing early detection and treatment-seeking behaviors. Both clinic- and community-based programs can cost-effectively deliver educational interventions to large numbers of individuals.

Evaluation of the effectiveness of clinic- and community-based programs for cancer control poses a number of methodological problems. The classical experimental design in which individuals are randomized into intervention and control groups frequently is not feasible. Even if it is possible to direct an educational program to specific individuals in a clinic or community, contamination through interaction between intervention and control participants, or diffusion of information from health care providers involved in the program to both intervention and control participants, can mask the true impact of a program.

One approach often taken is to identify two or more community settings and randomize these larger units to intervention and control conditions. However, this design results in several problems for program evaluation, most notably reduced statistical power for data analysis and increased complexity in estimating the needed sample size.
(Koepsell et al., 1992). The program evaluator must consider two kinds of sample size: (1) the number of communities randomized to each intervention and control group, and (2) the number of individuals sampled within each community. Koepsell et al. (Koepsell et al., 1992) note that for a fixed total number of individuals in a study, statistical power is almost always lower when allocation to intervention and control conditions is by community rather than by individual. Increasing the number of communities can increase statistical power, but can be prohibitively expensive.

Randomizing at the community level and conducting the evaluation using individuals within the communities as the unit of analysis also increases the complexity of the analysis. Murray and Wolfinger (Murray and Wolfinger, 1994) note that the observations for individuals within the larger assignment units will be correlated due to common experiences and/or selection factors. Under these circumstances, analysis of the outcomes for individuals can result in a badly inflated Type I error, resulting in a much greater chance of erroneously concluding that the intervention is effective. Murray and Wolfinger (Murray and Wolfinger, 1994) emphasize that when conducting the analysis at the individual level one must be able to include the assignment units as nested random effects in the analysis in order to achieve the desired Type I error rate.

An alternative approach to evaluation of community studies is to use a quasi-experimental design in which program outcomes are examined over time in the intervention unit (Cook and Campbell, 1979). If data can be collected at several points in time before and after introduction of the intervention, a trend analysis can be conducted for the outcomes of interest. The basic design employing trend analysis has been described as a simple interrupted time-series design (Cook and Campbell, 1979; Windsor, 1986). This design can be particularly effective if a fairly large number of measurement points for outcomes of interest are available and a sequential or piecewise regression methodology can be used to analyze the data. Piecewise regression can identify both sudden changes due to an intervention, as well as more gradual changes over time (Gillings et al., 1981).

The present study reports the use of piecewise regression to evaluate the effectiveness of a community cancer education program.

**Background**

The Forsyth County Cancer Screening Project (FoCaS) was a clinic- and community-based education program to increase screening for cervical cancer and breast cancer among low-income women. The primary objective of FoCaS was to increase breast and cervical screening among urban, low-income, African-American and Caucasian women age 40 or older. The intervention was a comprehensive education and support program targeted to women in low-income housing communities in Winston-Salem, North Carolina. A comparison group was selected from similar housing communities in Greensboro, North Carolina. The intervention program was conducted in the public health clinics and the community. The clinic-based program was directed by a project nurse who conducted educational programs for providers, identified eligible women attending the clinics, and provided the women education and assistance in obtaining mammography screening. The community program included direct education for individual women, small group workshops, distribution of printed educational materials and health fairs. A detailed description of the intervention program is presented in other publications (Tatum et al., 1997; Paskett et al., 1998).

Given the methodological issues involved in the evaluation of community studies with small numbers of communities, a multi-faceted evaluation approach was designed that included mixed cohort/cross-sectional samples surveyed before and after the program in the intervention and comparison counties, and a trend analysis of mammography screening in the public health clinic where program activities were based in the intervention county. The latter evaluation is reported in this paper.
Method

Sample
The Reynolds Health Center (RHC) located in Winston-Salem is a primary source of medical care for low-income individuals in Forsyth County, including women in the FoCaS study population. The adult population seen at RHC is about 60% white, 35% African-American, and 5% Hispanic and other races. Patient turnover is fairly large, particularly among younger adults receiving STD, family planning or prenatal care services. However, the target population of the FoCaS study, women age 40 or older, is the most stable segment of the population. During the study period, the number of visits by women 40 or older ranged from 13,921 in fiscal year 1991–1992 to 1208 in fiscal year 1994–1995.

The clinics as well as an in-house mammography facility operate on a sliding scale fee basis. All major components of the FoCaS program, both in the community and in RHC, encouraged participants to obtain screening at the RHC mammography facility. Thus, while it is possible that some women who responded to the program obtained mammograms at some other facility, trends in use of the RHC mammography facility over the duration of the project provided a good indication of program success regarding mammography screening.

Data for the evaluation of trends in screening at RHC were obtained from the weekly logs of all mammograms performed during the period of July 1991 through June 1995. Although the logs reported the age of the women, information on race of the women was not available from these logs, and separate analyses could not be conducted for white and African-American women.

Measurement

Mammography screening
The reports of mammograms performed were summarized in 2 month intervals. This interval length was selected to reduce instability due to small numbers and to adjust for gaps in the provision of screening services. Several times during the study period the mammography facility shut down for 1 or 2 weeks in a given month. The following month it would then operate at increased capacity to catch up with missed appointments. The use of 2 month intervals eliminated the misleading pattern of large drops and jumps in reported screening by month.

The data initially were examined in two ways: (1) the number of mammograms performed and (2) the number of mammograms per 100 visits by age-eligible women (40 or older). Both measures showed the same general pattern of increase when examined over time, but changes in the actual number of mammograms performed were partly confounded with variation in the size of the patient population during the study period. For this reason, the number of mammograms per 100 visits by age-eligible women was selected as the most appropriate outcome measure. The mean number of mammograms performed per 2 month period was 195 (sx = 42), ranging from a low of 121 for November–December 1991 to a high of 279 for May–June 1995. The mean base population for calculation of the rate was 2380 age-eligible women (sx = 279), ranging from a low of 1664 in July–August 1993 to a high of 2745 in March–April, 1994.

Preliminary examination of the changes in the mammography rate over time revealed a marked seasonal variation in mammography screening. Within each year examined, the screening rates were highest in May–June, with a smaller peak in September–October. The rates were consistently lowest during the holiday period in November–December. Twelve month moving averages were computed from the bi-monthly rates to eliminate the fluctuations due to seasonal variation. Rather than centering the averages to reflect the midpoint of the 12 month period as is typically done (Croxton et al., 1967), the rates were plotted to reflect the last 2 months of each 12 month period. This was done primarily for ease of interpretation to show the break point in the trend line when the program began and to
facilitate discussion of specific times when key components of the program became active.

Control for age

Currently, a tentative consensus appears to have been reached regarding the recommendation of annual mammographic screening for women age 40 and over (Mettlin and Smart, 1994; Ernster, 1997). However, throughout the duration of the FoCaS project the recommendations for screening women age 40–49 were inconsistent and controversial. For this reason, mammography screening rates were computed and analyzed separately for women age 40–49 and 50 or older.

Analysis

A piecewise or ‘segmented’ linear regression model was used to assess the effect of the intervention program on mammography rates (Neter et al., 1996). Fitting the statistical model requires a decision, preferably a priori, as to where each segment begins and ends relative to the total time period. These beginning and ending points are referred to as nodes. The FoCaS study had two major intervention components: (1) the in-reach component that provided educational services for both providers and patients at RHC, and (2) the community component that provided educational materials, one-on-one instruction and group instruction to women in the targeted housing developments. By June 1993, provider education in RHC had been implemented and the procedures for identifying and contacting eligible women were in place. Elements of the community education program also were underway by this time and the community program was completely implemented by December 1993. Therefore, we selected an a priori node for two segments of the line as the end of the period May–June, 1993. This point represents the break between preprogram baseline data and the beginning of the intervention (Fig. 1).

Examination of the bi-monthly data points by age revealed a possible second node at the end of the period represented by July–August, 1994. The mammography rates at this point in the program showed a distinct change for women in both age groups from the trend established after the program began. A second node was set at this point. It is important to emphasize that the second node was determined empirically rather than with regard to an anticipated change in program effects based on new program activities. Choice of this second node was supported based on sensitivity analyses which revealed negligible effects on the slopes in segments 2 and 3 when choosing nodes that were 2 points preceding or following the final node selected.

The analysis is based on a piecewise regression of 19 of the 2 month periods (data points) with three linear segments defined as the preprogram period, first program period, and second program period. The node between the preprogram and first program segments occurs at data point 7, and the node between the first and second program periods occurs at data point 14.

Ordinary least-squares estimation was employed to test for homogeneity of variance and autocorrelation. Testing for departure from the assumption of variance homogeneity was conducted using the method of White (White, 1980). This test revealed no significant departure ($\chi^2 = 15.9$, d.f. = 16, $P = 0.463$), indicating that the use of least-squares estimation was appropriate for this analysis.

The Durbin–Watson statistic was used to test for the presence of autocorrelation. In addition, a generalized Durbin–Watson statistic (which is a by-product of PROC AUTOREG included in the SAS statistical software) was used to assess the order of the autoregressive process for the error terms in the model (Vinod, 1973). Positive autocorrelation was present in both age groups. Thus, final model results include Yule–Walker (also referred to as Prais–Winsten) estimates of regression coefficients (SAS Institute, 1993). The adjusted estimates include the intercept and slopes for the three segments, and the first-order autocorrelation parameter corresponding to an autoregressive model of order one which was fit to the error structure.
Results

Examination of trends
The changes in mammography screening rates by age for the total period May–June 1992 to May–June 1995 are shown in Figure 1. Overall the trend lines suggest a different pattern of change for younger as compared to older women. The screening rates for women age 40–49 declined by about 0.7 mammograms per 100 visits during the preprogram period. The rates increased during the first program period and then leveled out between 8.5 and 9.0 per 100 during the second program period. In contrast, the rates for older women show an inconsistent pattern of small declines and increases during the preprogram period and through the first program period. However, a large and consistent increase in screening rates occurred during the second program period. The rates increased from 7.7 per 100 visits in July–August 1994 to 10.1 by the end of the second program period.

Regression analysis
The initial ordinary least-squares regressions of mammography rates on time revealed moderate but significant autocorrelation for both age groups. The Durbin–Watson statistic was 1.44 ($P = 0.001$) in the regression model for women age 40–49 and $0.855$ ($P < 0.001$) in the regression model for women age 50 or over. The models were corrected for serially correlated residuals using the previously described Yule–Walker estimator (SAS Institute, 1993). The autocorrelation parameters estimated by this procedure were $\rho = 0.351$ ($P = 0.182$) for women age 40–49 and $\rho = 0.459$ ($P = 0.074$) for women 50 or older. Although the autocorrelations were not significant, the significance of the Durbin–Watson statistic for both age groups prompted a reporting of results from autocorrelation-adjusted regression models. Adjusted model results are shown in Table I.

Overall, the results of the regression analyses support the conclusions drawn from visual inspection of the trend lines shown in Fig. 1. For women 40–49, no significant association exists between mammography screening rates and time in the preprogram period. The slope of the rates, as estimated by the regression coefficient ($b$), increases significantly during the first program period and levels off to a non-significant slope in the second program period. The analysis of changes
Table I. Interrupted time-series regression analysis of the mammography screening rates, adjusted for autocorrelation

<table>
<thead>
<tr>
<th></th>
<th>Regression of rate on time (N = 19)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b (Slope)</td>
</tr>
<tr>
<td>Age 40–49</td>
<td></td>
</tr>
<tr>
<td>significance of rate by</td>
<td>-0.076</td>
</tr>
<tr>
<td>time for each regression</td>
<td>first program period</td>
</tr>
<tr>
<td>segment</td>
<td>0.367</td>
</tr>
<tr>
<td>second program period</td>
<td>-0.034</td>
</tr>
<tr>
<td>significance of the</td>
<td>0.444</td>
</tr>
<tr>
<td>change in rate by time</td>
<td>first program period to first</td>
</tr>
<tr>
<td>across regression segments</td>
<td>program period</td>
</tr>
<tr>
<td></td>
<td>-0.402</td>
</tr>
<tr>
<td>Age 50 or older</td>
<td></td>
</tr>
<tr>
<td>significance of rate by</td>
<td>-0.014</td>
</tr>
<tr>
<td>time for each regression</td>
<td>first program period</td>
</tr>
<tr>
<td>segment</td>
<td>0.015</td>
</tr>
<tr>
<td>second program period</td>
<td>0.417</td>
</tr>
<tr>
<td>significance of the</td>
<td>0.029</td>
</tr>
<tr>
<td>change in rate by time</td>
<td>first program period to first</td>
</tr>
<tr>
<td>across regression segments</td>
<td>program period</td>
</tr>
<tr>
<td></td>
<td>0.402</td>
</tr>
</tbody>
</table>

in the slopes of the segments indicate that the slope of the mammography rates in the first program period is significantly larger than the preprogram slope. The slope of the second program period is significantly smaller than that of the first program period. The results suggest an initial program effect, with maintenance but no further increase of program effects in the second program period.

Among women 50 or older, the slope of the preprogram period shows no significant increase in mammography rates over time. The slope of the first program period is not significant, but the slope of the second program period shows a large and significant increase in the screening rates throughout this period. The change in slopes from the preprogram to the first program period is not significant, while the slope in the second program period is significantly larger than that of the first period. These data suggest a delayed effect of the program among older women.

Discussion

The results of the interrupted time-series analysis presented here suggest that the FoCaS project was effective in increasing mammography screening among women age 40 or older in our study population. The time-series data support the more traditional evaluation of the program’s effectiveness by means of the comparison of intervention and control communities. The first phase of the FoCaS evaluation, based on pre-test and post-test cross-sectional surveys, revealed a moderate intervention effect in terms of increasing mammography screening (Paskett, 1999).

The trend analysis also provides an additional dimension to evaluation of the program. The changes in mammography rates over time from preprogram baseline data to the first and second program periods suggest that the program had an immediate impact on younger women age 40–49 that eventually settled into a maintenance of the higher screening levels through the second program period. Older women showed no significant change in screening rates early in the program, but their rates increased sharply during the second program period.

We did not anticipate this apparent differential program effect and cannot clearly attribute the differences by age to specific program components or to the timing of program initiatives. The in-
reach program in RHC began earlier than the community education program. Younger women may have been more readily accessible and/or responsive to being contacted by the project nurse while a patient at RHC, resulting in an earlier increase in their screening rates. The fact that the screening rate for women age 40–49 did not continue to increase throughout the second program period may be partly due to the mixed messages women received regarding the value of mammography screening for younger women. As mentioned earlier, the controversy existed throughout the FoCaS project. Midway through the project, the National Cancer Institute changed its stance from recommending screening for women age 40–49 to not recommending screening for women in this age group. The return to the original recommendation of screening occurred after the intervention program had ended. If this controversy did affect women in our study population, it may be significant that the project was able to avoid a decline in mammography screening rates for women 40–49 in the second program period.

While women age 50 or older received consistent messages regarding the value of screening throughout the project period, increasing cancer screening typically is more difficult among older women. The community education program was not fully implemented until several months after the in-reach program began at RHC. The community program was very well received by older women and it may be that the combined effects of both components, in-reach and community education, were needed to eventually have an impact upon older women. Thus, a delayed effect emerged in the trend line.

One additional factor must be considered in interpreting the results for both older and younger women. The major threat to the internal validity of a single-group (intervention community) time-series design is history (Cook and Campbell, 1979; Windsor, 1986). That is, unplanned events or exposures that may occur at the program site which impact upon the outcome(s) of interest. The North Carolina Breast and Cervical Cancer Control Program (BCCCP), funded by the Centers for Disease Control, was in operation throughout the FoCaS project period. This program offers free screening to women through the public health department, and the target populations of BCCCP and FoCaS are similar. In fact, FoCaS worked in concert with BCCCP during the project period by identifying BCCCP as a possible source of free mammograms for eligible women participating in the FoCaS project.

An important point in support of an independent effect of FoCaS is that BCCCP was active throughout the FoCaS preprogram period, with no evidence of an increase in the mammography screening rates during this period. However, BCCCP began a special initiative called ‘Woman-Wise’ shortly after the FoCaS intervention began that may have increased its impact.

Other qualitative information supporting an independent effect of FoCaS lies in the nature of the BCCCP program. This program is funded through the North Carolina Department of Health and Human Services. While BCCCP is available to all income-eligible women 40 or older, the state instructs and provides incentives to county health departments to direct the program to women over 50. In Forsyth County, only about 16% of the approximately 350 provided mammograms yearly by BCCCP are 40–49. Thus, the possible confounding effects of BCCCP and FoCaS likely are weaker for women age 40–49.

Finally, the BCCCP program also was active in the comparison community during the study period and the analysis of changes in reported mammography screening for the cross-sectional survey samples revealed a significantly greater increase in screening in the intervention community. The consistency of findings for the traditional intervention–control evaluation and the trend analysis reported here strengthen the overall conclusion of an intervention effect.

However, the inability to completely separate the effects of FoCaS and BCCCP is a weakness of this analysis and represents a potential problem in all community-based studies that use a quasi-experimental design, regardless of whether the design is time-series or intervention–control. An
unrelated program with similar objectives conducted in the control community and not in the intervention community can mask study intervention effects. Conversely, the same program implemented in the intervention but not the control community can be misleading regarding the strength of intervention effects. Assessing history as a threat to internal validity in community-based studies using a quasi-experimental design frequently will require both quantitative and qualitative analysis to come to a reasonable conclusion about the effects of the study intervention.

Other weaknesses in the present study include (1) a lack of information on the trend of mammography rates after the FoCaS program ended and (2) a comparatively short time-series (19 data points) used for the analysis. Regarding the first point, ideally a full interrupted time-series design will measure several points before, during and after the intervention to determine what occurs after it ends (Windsor, 1986). Secondly, formal time-series designs generally require 50 or more data points to accurately determine the structure of the correlated error in the data (Cook and Campbell, 1979; Windsor, 1986). However, Cook and Campbell (Cook and Campbell, 1979) point out that even if formal statistical analysis is limited or not possible, visual inspection of a short time-series together with an analysis of threats to validity can provide useful information on the effectiveness of an intervention.

In general, interrupted time-series designs will be most useful if longitudinal data on the primary outcome(s) of interest are available as routinely collected data or if the desired information can be obtained fairly inexpensively prior to, during and after the program. In the context of community health education, examples include cancer screening rates, birth outcomes (e.g. preterm birth, low birth weight, infant mortality, prenatal complications), vaccination rates and use of STD or family planning service.

Although not yet widely used in health education, studies which have reported successful use of regression analysis to analyze single-group interrupted time-series designs include an evaluation of alcohol treatment programs (Berman et al., 1984), assessment of the effects of changes in Medicaid reimbursement procedures on nursing home costs (Coburn et al., 1993) and evaluation of a program to improve the diet of elementary school students (Coates et al., 1981). Other studies have used an interrupted time-series design with non-equivalent control groups to evaluate the effectiveness of a perinatal care program in North Carolina (Gillings et al., 1981) and a program to increase the participation of volunteers in hospital care for elderly patients (Laitinen et al., 1996). The addition of a control group to an interrupted time-series design results in a particularly strong design that allows a control for history if the intervention and control groups are exposed to the same non-program influences.

Introduction of an interrupted time-series evaluation component can strengthen a study that is based on the frequently used quasi-experimental design in health education research which assigns intervention and control status at the community or organizational level, and examines outcomes at the individual level before and after conducting the intervention. If longitudinal data on program outcomes are available in the intervention community, these data will provide an independent test of the intervention effect that can support the overall conclusions regarding the success of the program. The time-series analysis also will provide information on the timing and maintenance of intervention effects for all participants and for different subgroups in the study population. This information is particularly important for programs that have two or more intervention components introduced at different times. If longitudinal data on one or more intervention outcomes can be collected in both the intervention and control communities, the resulting evaluation will be even stronger.

The single group interrupted time-series design also can serve as an effective alternative to a traditional intervention–control design in situations where funding limitations allow introduction of a program in only one community, social policy dictates that all eligible communities receive the
program or one wishes to evaluate a natural occurring social intervention such as a major change in the health care delivery system in a community. Under these circumstances the interrupted time-series design is much stronger than a single group before-after evaluation, in which it is not possible to determine if any changes in the outcome of interest are due to a genuine intervention effect or simply a continuation of past trends.

The present study demonstrates the utility of the interrupted time-series quasi-experimental research design. This approach should be considered as a viable option for program evaluation whenever true randomization to intervention and control groups is not possible.

Acknowledgement

This research was supported by a grant from the National Cancer Institute (CA57016).

References


Received on May 13, 1999; accepted on February 2, 2000