Adding a Time-Series Design Element to the Success Case Method to Improve Methodological Rigor
An Application for Nonprofit Program Evaluation

Chris L. S. Coryn
*Western Michigan University*
Daniela C. Schröter
*Nexus Market Research, Inc.*
Carl E. Hanssen
*Hanssen Consulting, LLC*

**Abstract:** Brinkerhoff’s Success Case Method (SCM) was developed with the specific purpose of assessing the impact of organizational interventions (e.g., training and coaching) on business goals by analyzing extreme groups using case study techniques and storytelling. As an efficient and cost-effective method of evaluative inquiry, SCM is attractive in other contexts as well, although few examples of such uses are to be found in the published literature. However, modifications of the SCM concept and design are sometimes necessary for implementing the approach in nonprofit environments where business goals are not necessarily an explicit objective. This method note demonstrates how SCM was modified and extended to a social service context, in which the program evaluated was aimed at reducing chronic homelessness and unemployment. Modifications included defining success in a nonprofit setting and adding a time-series element to the design features of traditional SCM to increase methodological rigor.

**Keywords:** success case method; case studies; outcome evaluation; impact evaluation; evaluation methodology; time-series

Brinkerhoff (2003) developed the Success Case Method (SCM) to evaluate the impact of interventions on business goals. In general, SCM is conducted with the intent of assessing how well an organizational intervention is working (primarily in for-profit contexts) by focusing on extreme groups and identifying the contextual factors that differentiate successful from unsuccessful adopters of new initiatives. Although SCM has been applied primarily in the for-profit sector to evaluate training initiatives or new work methods, the technique has also been...
used for evaluating the impact of a U.S.-based and international food security project (Chianca & Risley, 2005, October), in educational settings to determine the factors that influence the academic achievement of minority students (Coryn et al., 2007) and to evaluate organizational learning in a nonprofit foundation (Berkley, Day, Smith, & Chianca, 2005, October). More recently, SCM has been proposed as an alternative approach for examining causal associations when more scientifically rigorous, sophisticated, and elegant designs are unethical, unpractical, too costly, or simply unfeasible (Brinkerhoff, 2005; Davidson, 2005; Scriven, 2006a, 2006b). Traditionally, SCM is conducted using a five-step procedure:

1. Focus and plan the SCM
2. Create an impact model
3. Survey all program recipients to identify success and nonsuccess cases
4. Interview a random sample of success and nonsuccess cases and document their stories
5. Communicate findings, conclusions, and recommendations

The focus and planning of a SCM study can take many forms and can be used for both formative and summative purposes (Step 1). Once the focus of the study has been determined, an impact model is developed that delineates how an intervention is assumed to produce its desired results (Step 2). Then, cases are identified as high (H), that is, success cases; moderate (M), or average cases; or low (L), that is, failure or nonsuccess cases (Step 3). Typically cases are classified using survey methods specifically designed to determine the classification of cases. Once classified, these cases serve as sampling strata for Step 4. SCM is, therefore, essentially an analysis of extreme or outlier cases, as opposed to average cases, where independent evidence is sought to corroborate claims of success or failure (e.g., that a salesperson’s sales actually increased following training). Next, the reasons underlying successes or failures are investigated using semistructured interview techniques designed to probe possible explanations from a random sample of extreme cases (Step 4). Finally, the SCM findings, conclusions, and recommendations are communicated (Step 5). Often, reports are presented in the form of “success stories.” Figure 1 illustrates the basic logic in applying SCM (where an observable effect has occurred and cases can be accurately classified as high, medium, or low, in terms of the magnitude of the desired effect).

In the remainder of this method note, we illustrate how traditional SCM methodology was modified for evaluating the impact of programmatic interventions aimed at improving the lives of a marginalized social group. Typically, SCM studies are used to estimate business goals such as return on investment (ROI; a performance measure used to evaluate the efficiency of an investment or to compare the efficiency of a number of different investments. To calculate ROI, the benefit, or return, of an investment is divided by the cost of the investment, and the result is expressed as a percentage or a ratio). Given the nature of the evaluand described in the case example, however, ROI would not have been the most meaningful measure of success.

The Case Example

Overview of the Initiative

The case example of our modification of Brinkerhoff’s SCM is taken from an evaluation of a local-level program aimed at reducing and preventing homelessness and chronic unemployment. The program was developed in response to a well-documented need in a mid-sized Midwestern city, with the long-term objective of reducing recipients’ need for public
assistance. The program was established as a joint effort between three local service providers who initially offered independent programs to the same target group but under rules and regulations established by three different funding sources with different outcome goals. The collaborative program provided pooled services to recipients for 1 year in an effort to stabilize housing and employment needs. One specific feature of the program was that services were tailored to individual needs via intensive case management, housing subsidies, and other unique service provisions (e.g., budget management workshops, time management workshops, interviewing skills training). Recipients’ needs were reassessed monthly in a meeting of case managers from each of the three collaborative agencies. In addition to services and support provided through the program, participants were referred to other programs for service needs not covered by the collaborative. The program served approximately 75 individuals and families per annum and as a relatively young program (established in 2001) intended to expand its services to a larger target group, enhance its linkages to partners, improve its information systems, and affect policies impeding efficient, effective collaboration across community agencies. The program also engaged in numerous other supportive services and activities, including wrap-around case management, monthly team meetings, and public policy advocacy.

SCM, as used within this example, was implemented as one component of a larger, comprehensive evaluation, conducted by Hanssen Consulting, LLC, using Stufflebeam’s (2002) context, input, process, and product (CIPP) evaluation model, of which the modified SCM was fit within the effectiveness evaluation component of the CIPP evaluation (i.e., product). The
primary rationale for using SCM to assess the program’s impact on program recipients was to plausibly eliminate rival hypotheses about core factors leading to sustainable success of service recipients and to do so in an efficient and cost-effective manner. For example, what program elements did successful program recipients embrace and utilize? Did recipients utilize aspects or elements of the program differentially? What factors supported or inhibited program outcomes and impacts? Or, did outcomes, impacts, and effects occur from unknown origins, sporadically, or were they idiopathic? Simultaneously, a longitudinal quasi-experimental study designed to compare differences in anticipated outcomes between program recipients and a statistically matched control group was also conducted to strengthen any causal inferences produced from the modified SCM. Although we employed a mixed methods approach for the larger evaluation to build a base of evidence for any claims regarding the program’s effectiveness by triangulating the results produced by each independent method (Greene, 2007), we also propose that this modification of the SCM can be used as a stand-alone approach, albeit one that is largely qualitative (and normally using small samples).

Our modifications to the SCM were twofold and intended to serve multiple purposes. These included, but were not limited to, (a) defining success in a context which did not have an observable, measurable ROI or impact on the service provider and (b) adding a longitudinal, time-series design element to traditional SCM methodology to increase methodological rigor.

Defining Success in a Nonprofit Context

Our case scenario required redefining the meaning of success typically associated with SCM studies, which are usually the impact of an intervention on organizational results such as ROI. Although it is certainly possible to determine a nonmonetary ROI or cost-benefits ratio (Levin & McEwan, 2001; Yates, 1996), this may not be the most meaningful measure of success in cases like this, where success is somewhat removed from an organizational setting and instead applies more directly to a personal and social one. Our definition of success required rethinking and reconceptualizing what is meaningful, not only for the collaborative service providers but also in terms of its recipients and downstream impactees. In essence, the meaning of success in this context could be considered across three levels: (a) upstream stakeholders (e.g., program service providers, designers, funders), (b) immediate impactees/consumers (e.g., the program recipients and their families), and (c) downstream impactees/consumers (e.g., the larger community).

Based on feedback from these three stakeholder groups, our conceptual definition of success for the case scenario presented here is threefold: (a) long-term retention of economically self-sufficient employment, (b) stability of housing, and (c) reduced need for government or other social support systems (e.g., TANF, food stamps). Combined, these three objectives then became the criteria used to identify and classify success and failure cases, where a “success” case would have positive outcomes across all three objectives, an “average” case would have positive outcomes on one, or perhaps two, of these objectives, and a “failure” case would demonstrate poor outcomes across all three.

In essence, success or failure of the overall program was determined by any social and economic benefits (or detriments) that occurred both directly and indirectly as a result of its activities. In addition, program success or failure was also gauged to the extent that meaningful benefits for program recipients (e.g., sustainable, self-sufficient income and improved quality of life) and, therefore, also the immediate community (e.g., reduced consumption of public monies and services) rather than the program itself, were produced.
Increasing Methodological Rigor by Adding a Time-Series Design Element to the Success Case Method

Figure 2 presents the general logic and design elements of the modification that we made to Brinkerhoff’s SCM for use in the context of a human/social service program using the definition of success outlined above (again, we reiterate that this modified design could be applied equally well in other settings). As can be seen in the figure, we simply added a time-series component (represented by $t_1$, $t_2$, and $t_3$) to the traditional SCM. As we intend to demonstrate, this modification offers a number of practical advantages for applying SCM in the human/social service context, although only moderately increasing the complexity, time, and resources associated with traditional SCM methodology.

Some Advantages of the Modified Success Case Method

Whereas Figure 2 illustrates our modification for applying the modified SCM in human/social service contexts, Figure 3 illustrates some of the advantages offered by adding a longitudinal, time-series design element to traditional SCM, which include (a) the ability to identify growth (dashed, upward turning arrows in Figure 3) and decay (dashed, downward turning arrows in Figure 3) and the reasons for them (e.g., delayed manifestation of an effect following a treatment), (b) the ability to identify long-term program effects and for who and why (or why not) those effects are, or are not, sustained, and (c) the ability to provide useful feedback to the program at various points during the evaluation (represented by the feedback loops from $t_1$, $t_2$, and $t_3$ to the program in Figure 3).
While SCM methodology is not intended to generalize to the average participant (Brinkerhoff, 2003), the addition of the time-series design element was also intended to reduce some of the threats to internal validity inherent in most single-group designs by identifying and eliminating as many plausible, competing explanations for observed effects as possible. Therefore, we would also assert that another benefit of the modified SCM is that by increasing methodological rigor through the addition of design elements, causal inferences can be better supported when stronger cause probing designs (e.g., randomized experimental designs, regression discontinuity designs, interrupted time series designs) are not feasible.

**Implementation of the Modified Success Case Method**

**Sample Characteristics**

At $t_1$, success (H) and failure (L) cases that completed the 1-year program were identified (6 months after program completion) following traditional SCM procedures and were classified based on the three criteria outlined earlier. Of the ~75 individuals who had completed the program 6 months earlier, approximately equal proportions were classified as H (32%), M (40%), and L (28%) cases and served as the sampling frame for the remainder of the study. Largely, these individuals were single, African American women, who had more than one child, were between the ages of 27 and 37, and had obtained less than a high school education or held a GED (see Table 1). Nearly all were receiving public assistance in the form of food stamps and...
cash payments (i.e., TANF) prior to participating in the program and a large proportion were either doubled-up (i.e., living with someone) or staying in temporary shelters prior to the intervention.

Interview participants were compensated with a US$10 voucher for gasoline and a US$10 voucher for a local supermarket (participants not owning automobiles received US$20 for a local supermarket). Baseline, preprogram data on income, employment, housing, and amount and types of public assistance, among others, were gathered when individuals applied to the program to better support any claims that improvement (i.e., “success”) could be attributed to the intervention, and these same data were collected again from all participants in the sample frame at the three, equally spaced time intervals.

A random sample of approximately one third of the two extreme strata (i.e., H and L cases) was interviewed at $t_1$ using a semistructured interview protocol designed to investigate reasons for success or failure (Brinkerhoff, 2003) and to probe for alternative explanations of success or failure not directly attributable to the program. Generally, these interviews emphasized various aspects of employment, housing, and public assistance prior to, during, and after participation in the program (and, to some extent, these interviews incorporated event-history techniques). The same procedure was followed at $t_2$ (12 months after program completion) and $t_3$ (13 months after program completion), although the composition of the H and L strata changed slightly as individuals moved from one to another on reclassification according to the success criteria (e.g., Did an individual classified as H at $t_1$ move to a new classification, M or L, due to job loss at $t_2$?). At both $t_2$ and $t_3$, the original randomly selected cases were reinterviewed and newly classified cases (e.g., cases where an individual moved from M to H or M to L; that is, growth and decay, sustainability and durability of effects) were added to the sample to investigate these changes in status and the timing of program treatment effects. The time intervals of the design were selected on the basis of extra-study knowledge (e.g., from the experiences of program staff, scientific literature) about when such effects are likely

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant Characteristics</td>
</tr>
<tr>
<td>Characteristic</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Marital status</td>
</tr>
<tr>
<td>Single</td>
</tr>
<tr>
<td>Married</td>
</tr>
<tr>
<td>Education</td>
</tr>
<tr>
<td>Less than high school</td>
</tr>
<tr>
<td>High school or GED</td>
</tr>
<tr>
<td>Some college</td>
</tr>
<tr>
<td>College degree</td>
</tr>
<tr>
<td>Race/Ethnicity</td>
</tr>
<tr>
<td>African American/Black</td>
</tr>
<tr>
<td>Caucasian/White</td>
</tr>
<tr>
<td>Other (e.g., Latino/Hispanic)</td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Number of children</td>
</tr>
<tr>
<td>Food stamps</td>
</tr>
<tr>
<td>TANF</td>
</tr>
</tbody>
</table>

Note: GED = general educational development; TANF = temporary assistance for needy families.
to occur as well as their probable magnitude. Making the timing of observation intervals equal is an important feature of this design because it permits assessment of linear changes over time, and comparisons between \( t_1 \) and \( t_2 \) and between \( t_2 \) and \( t_3 \) would be less meaningful if the time intervals differed.

**H, M, and L Growth, Decay, and Outcomes Across Time**

Using the modified SCM, we were able to identify and map growth and decay across time (i.e., sustainability and durability of effects or differential onset of treatment effects across strata). As shown in Figure 4, strata composition at the three times points were \( H = 32\% \), \( M = 40\% \), and \( L = 28\% \) at \( t_1 \); \( H = 38\% \), \( M = 38\% \), and \( L = 24\% \) at \( t_2 \); and \( L = 40\% \), \( M = 30\% \), and \( L = 30\% \) at \( t_3 \). Although movement between strata was relatively small, with total growth of 18% and total decay of 12% across all three time periods, careful study of the two extreme strata revealed that the almost nonexistent upward growth from the L to M strata across time (4% at \( t_1 \) and 0% at \( t_2 \)) was, in fact, due to a dose–response relationship. That is, participants initially classified as L at \( t_1 \) received lower “dosages” of the intervention than those classified as H or M. Those classified as L often missed many of their monthly case meetings or failed to complete time management or budgeting workshops and other development opportunities and activities, for example. Conversely, those classified as H received higher “dosages” of the intervention. That is, they regularly attended their monthly case meetings and completed or participated in all of the services and activities offered by the initiative. In no case did a participant cross two strata (i.e., from L to M to H or from H to M to L).

These simple examples, explaining differential outcome patterns and dose-response relationships, provided useful feedback to the initiative (see feedback loops in Figure 3) and were used to refine service delivery and participant selection practices and procedures with the intent of improving outcomes for future participants. In addition, this design feature also
provided useful information regarding outcome permanence (i.e., continuous or discontinuous) and immediacy (i.e., immediate or delayed) than would have been the case with traditional SCM.

During the 18-month period studied, participants classified as H received increases in their hourly wages, number of hours worked, and as a result, were able to maintain stable, affordable housing (on average, annualized housing costs to annualized income was 65% at baseline and by \(t_3\) was 48%). While receipt of food stamps remained relatively stable (\(M = \text{US}\$246\) per month) for this and the other two strata, they were able to significantly reduce their reliance on TANF (from \(M = \text{US}\$268\) per month at baseline to \(M = \text{US}\$147\) per month by \(t_3\); partly attributable to increased income and stable employment). Those classified as H experienced even greater benefits than the average of all program participants and those classified as L frequently experienced more negative outcomes (e.g., unemployment, unstable housing, increased need for and reliance on public assistance). On average, across all three strata combined, the slopes for employment-related outcomes (e.g., income) were positive across time (i.e., increasing), whereas the slope for amounts of public assistance received was downward across time (i.e., decreasing). The slope for income-related outcomes was steeper than for public assistance outcomes, however, indicating that the initiative was slightly more successful at improving income-related outcomes than at reducing reliance on public assistance. Although reliance on public assistance declined only moderately across time, the savings to the public in terms of resource usage (e.g., consumptions of public monies and services) were, on average, in the thousands of dollars on a monthly basis across all of the initiative’s participants combined, whether individual participants themselves were classified as successful or not. All in all, these data suggested that the initiative was successful in the way that success was defined for the study.

**Can Causal Inferences be Made From the Modified SCM?**

At the core of Brinkerhoff’s traditional SCM is determining whether outcomes or effects can be attributed to an intervention or treatment. However, demonstrating that an observed pattern of outcomes or effects are causally associated with an intervention or treatment without a source of counterfactual inference can be problematic unless one has extremely complex causal hypotheses, thus making alternative explanations less plausible. By adding structural design elements, we sought to reduce the number and plausibility of internal validity threats associated with traditional SCM (e.g., all of the threats relevant to one-group posttest-only designs; see Campbell & Stanley, 1963; Shadish, Cook, & Campbell, 2002) as a method of descriptive causal investigation.

Much like a quasi-experimental time-series design or interrupted time-series design, our variant of SCM allows control over some threats to internal validity, making a more compelling argument for using SCM as a reasonably rigorous, reliable, credible, and viable alternative for making some types of causal inferences (e.g., those not necessarily requiring standard scientific precision but rather of a degree of certainty by which decision makers can render reasonable judgments and make sound, informed decisions about the general disposition of a program or ways to improve it; c.f., Davidson, 2005, 2006, 2007; Stallworth & Roberts-Gray, 1987), particularly those aimed at analytic generalization versus statistical regularities (Scriven, 2008; Yin, 1994).

However, modifying design elements alone do not sufficiently rule out alternative explanations for observed effects in single-group studies. Only making alternative explanations explicit and ruling them out one-by-one to render them as being unlikely causes will support such
claims. In large part, our ability to make inferences about cause–effect associations was supported by the context-bound, collective evidence that supported these conclusions, which were produced by systematically ruling out various threats to validity and likely alternative explanations for observed effects. Methodologically, this was accomplished in numerous ways, including negative-case sampling (i.e., through extensive analysis of L cases) where cases disconfirmed the expected pattern of results, causal tracing using modus operandi method (MOM) and general elimination method (GEM) techniques to confirm the timing of observed outcomes and their sequential ordering (Davidson, 2000, 2005; Mohr, 1999; Patton, 2008; Scriven, 1976, 2008), and using low-inference descriptors (i.e., participants’ descriptive accounts of the timing of results) to match outcome patterns and produce simple causal descriptions.

In the Shadish et al. (2002) typology, there are nine specific threats to internal validity that, if ignored, suggest that a suspected cause–effect relationship might not be in fact causal, would have occurred in the absence of a treatment, or may be due to other possible causes. In the case presented here, some threats were more plausible than others. Two of the most serious internal validity threats were maturation and regression (although there were others, these two will serve as explicit examples). Maturation threats are those things that can or do occur naturally in the absence of a treatment or intervention, such as cognitive development, aging, and learning. These types of threats are problematic when they could produce the effect attributed to the treatment or intervention. Particularly relevant here are secular trends (Rossi, Freeman, & Lipsey, 1999), such as growth in the local economy, which might account for employment outcomes, for example. However, local secular trends and events that might influence or explain changes (e.g., employment trends, opening or closing of industries in which participants worked) were closely monitored and could, therefore, be ruled out as an alternative explanation for observed employment outcomes.

Regression is a very real internal validity threat for SCM studies, which overtly examine extreme cases at the upper and lower ends of a distribution. This particular threat was addressed in our modified version by adding additional multiple posttest observations to the design. By adding additional observations, we were able to explicitly explore patterns of growth and decay at an individual level and, therefore, determine whether extreme cases (i.e., H and L) were “true” extreme cases. Instrumentation, on the other hand, which is particularly salient in longitudinal studies and is best avoided by consistency of measurement devices (i.e., not changing devices or meaning of variables of interest midway through a study), was not a serious threat. This is true because even though our interview protocols were semistructured, they were facilitated by the same two evaluators (working together) at all three time points and who shared a common meaning of the constructs being measured.

In identifying and ruling out such threats, numerous competing explanations (i.e., rival hypotheses) for observed treatment effects are identified and can, in some cases, be eliminated as plausible alternative explanations. This does not, however, eliminate all potential rival or competing explanations, which is better accomplished by a thorough analysis of the local context, extensive extrastudy knowledge, and careful questioning of those who benefited (or failed to benefit) from the treatment or intervention, all of which were given extensive attention in this particular study.

What could not be systematically ruled out was the fact that some participants were also receiving similar services from other local organizations (i.e., history). A history threat is an event other than the treatment or intervention that could have produced the observed outcome in the absence of the treatment or intervention. Thus, were the observed outcomes attributable to the intervention of interest, to other local interventions with similar objectives, or to an interaction? Without an equivalent comparison group, such threats are difficult to eliminate.
The balance of the evidence, however, supported our conclusions that improved employment- and housing-related outcomes for participants, as well as their reduced need for public assistance, could be causally attributed to the initiative.

**Conclusions**

Although we are encouraged by what we see as the strengths that our modified SCM contributes to traditional SCM and single-group designs in general, there still remain numerous limitations, most of which could be reduced by adding additional, carefully chosen design features. Multiple pretests, for example, would serve to even further diminish maturation and regression threats by identifying any potential biases operating prior to the introduction of the intervention (e.g., that income trajectories were already rising, that unemployment was temporary). The addition of a multiple pretest design element would enhance the modified SCM even further by more closely approximating some of the characteristics of short interrupted time-series designs (see Shadish et al., 2002). In addition, measurement of one or more nonequivalent dependent variables (a dependent variable that is not expected to change because of the treatment or intervention but is expected to respond to contextually important internal validity threats in the same way as the target outcome) would strengthen the design even further, with relatively little increase in time or costs.

Even though SCM is an efficient and cost-effective method of assessing the impact of organizational interventions (which is part of the reason why it has garnered such acclaim in corporate settings), our modification for applying the method in the context of human and social service (i.e., nonprofit) program evaluation does reduce both of these advantages to some extent by adding to the time and resources required to conduct the time-series SCM. However, these limitations are offset by the gains in understanding of the onset (i.e., whether immediate or delayed) and permanence (i.e., whether continuous or discontinuous) of effects (as well as greater insight into who the program works for, who it does not, and why, by analysis of subsets or units across time) that would otherwise not be detected with traditional SCM.

SCM, and our modified version of it, is best suited for interventions targeted at small to medium populations (e.g., local- or community-level interventions), where alternative explanations for an observed effect can be thoroughly probed and internal validity threats of causal claims about whether an interventions “works” or actually produced an observed effect can be identified and ruled out. Therefore, and even though we assert that the modified SCM can, in certain situations and meeting certain conditions, be used to make reasonable claims about cause–effect relationships, we do not believe that it would meet the standards required for providing such evidence for large-scale, national-level interventions or programs (for a complete discussion of the contemporary debates about evidence-based evaluation methodologies, see Donaldson & Christie, 2005; Julnes & Rog, 2007). Moreover, this type of design examines only a small subset of descriptive causes and fails to account for the many causal forces that operate simultaneously and in complex, nonlinear ways.

In addition to the limitations associated with the modified SCM in general, the implementation demonstrated in the case example also suffered from more and less serious limitations. Most notable among these was that our ability to build strong causal inferences was confined within the limits of our capacity to enumerate specific contextual threats. As with all such cases, we have yet to develop methods for quantifying qualitative threats. Therefore, we easily could have overlooked an important, plausible alternative explanation for the effects that were observed. People are just not good at making causal inferences and they are fallible (e.g., confirmation bias). Also, we suspect that patterns of growth and decay (i.e., onset and
permanence of effects) in other implementations will likely not be as easily interpreted as those observed in our case example due to more complex patterns. In addition, the population of participants for our case example was rather small. Such might not be the case in future implementations where postprogram attrition or time-series follow-up observations might cause both methodological and analytic problems, particularly given the method’s emphasis on analysis of extremes of the distribution (i.e., H and L cases). Finally, we did not explore potential factors that might mediate or moderate outcome patterns, such as motivation or prior employment history, which would more fully explain how, why, and for whom the initiative worked or did not work. More complex causal hypotheses, such as those including mediating variables, would provide a more compelling case for cause–effect associations attributable to the initiative.

Although SCM has been most widely used to estimate the impact of individuals’ successes and failures on organizational goals (e.g., ROI), its use in nonprofit contexts requires an expansion from organizational goals to other groups of stakeholders. As organizations whose justification solely lies in meeting consumers’ needs (Coryn, Gugiu, Davidson, & Schröter, 2008), the success of nonprofit programs such as the one described here must be valuable not only to the people it serves but also to their immediate families and other social networks as well as the larger community. The benefits from such a program should not only be measured in terms of organizational goals and objectives but also be seen in relation to consumer and community need.

In addition to reconsidering definitions of success in a nonbusiness context (although the modified method could be used in such contexts), the case has shown that a modification of the traditional, one-group posttest-only SCM to a time-series SCM (with additional design considerations) can decrease some, but certainly not all, threats to alternative explanations for observed patterns of effects. Although modifying a study’s design features, making all possible threats explicit, and then ruling them out do alleviate some threats to the validity of causal inferences, they do not, however, ameliorate all of them.

For evaluators working in nonprofit contexts and with little or no access to equivalent (and even nonequivalent) comparison groups, SCM can serve as a useful evaluation method regardless if it is used as a stand-alone technique or as a supplement to others. The time-series SCM presented in this article can provide useful information and important findings about a program’s impact on consumers and larger communities not possible with traditional SCM. Although our modified SCM will certainly be viewed with skepticism by those engaged in contemporary dialogues about evidence and methods-choice, it does, nonetheless, answer important evaluative questions and provide plausible evidence as to whether a program is working, as well as having the potential to be used both formatively and summatively.

References


