Integrating a Comparison Group Design into a Theory of Change Evaluation: The Case of the Urban Health Initiative
Beth C. Weitzman, Diana Silver and Keri-Nicole Dillman
American Journal of Evaluation 2002; 23; 371
DOI: 10.1177/109821400202300402

The online version of this article can be found at: http://aje.sagepub.com/cgi/content/abstract/23/4/371
Articles should deal with topics applicable to the broad field of program evaluation. Articles may focus on evaluation methods, theory, practice, or findings. In all cases, implications for practicing evaluators should be clearly identified. Examples of contributions include, but are not limited to, reviews of new developments in evaluation, description of a current evaluation study, critical reviews of some area of evaluation practice, and presentations of important new techniques. Manuscripts should follow APA format for references and style. Length per se is not a criterion in evaluating submissions.

Integrating a Comparison Group Design into a Theory of Change Evaluation: The Case of the Urban Health Initiative

BETH C. WEITZMAN, DIANA SILVER, AND KERI-NICOLE DILLMAN

ABSTRACT

This paper describes how we strengthened the theory of change approach to evaluating a complex social initiative by integrating it with a quasi-experimental, comparison group design. We also demonstrate the plausibility of selecting a credible comparison group through the use of cluster analysis, and describe our work in validating that analysis with additional measures. The integrated evaluation design relies on two points of comparison: (1) program theory to program experience; and (2) program cities to comparison cities. We describe how we are using this integrated design to evaluate the Robert Wood Johnson Foundation’s Urban Health Initiative, an effort that aims to improve health and safety outcomes for children and youth in five distressed urban areas through a process of citywide, multi-sector planning and changed public and private systems. We also discuss how the use of two research frameworks and multiple methods can enrich our ability to test underlying assumptions and evaluate overall program effects. Using this integrated approach has provided evidence that the earliest phases of this initiative are unfolding as the theory would predict, and that the comparison cities are not undergoing a similar experience to those in UHI. Despite many remaining limitations, this integrated evaluation can...
INTRODUCTION

This paper describes how we strengthened the theory of change approach to evaluating a complex social initiative by integrating it with a quasi-experimental, comparison group design. The Urban Health Initiative (UHI), sponsored by the Robert Wood Johnson Foundation (RWJF), aims to improve the health and safety of children and youth in distressed urban areas. Several features of the initiative, including its non-prescriptive guidelines, its focus on entire cities, and its emphasis on changing systems, made it difficult to identify an appropriate evaluation design.

We begin this paper by discussing the contributions and limitations of the theory of change approach in helping evaluators assess impacts, and review prior suggestions for integrating such an approach with a more traditional quasi-experimental design. Next, we describe our case example and demonstrate how the use of these two research frameworks and multiple methods is enriching our ability to test underlying assumptions and evaluate overall program effects. We also demonstrate the plausibility of selecting and validating a credible comparison group through the use of cluster analysis.

Despite the initiative’s complexity, the evaluation’s design allows us to better answer the all-too-familiar question in evaluation research, “Does the program work?” “Does the program work?” To answer this question, one must compare observed outcomes with what “would have happened if the world had been exactly the same as it was except that the program had not been implemented” (Hatry, Winnie, & Fisk, 1981, p. 25). While random assignment may permit researchers to most accurately estimate the counterfactual, quasi-experimental designs, such as those with carefully selected comparison groups and repeated measures, permit researchers to estimate it credibly as well. Ultimately, only a well-crafted evaluation design with a legitimate point of comparison can support conclusions that observed impacts are in fact the result of the intervention, rather than alternative explanations.

Yet many innovative social policy ideas do not readily lend themselves to traditional evaluation designs. Well-defined and well-understood interventions are not the “order of the day” for much social policy, as practitioners and funders, seeing a complex world, have responded with increasingly complex interventions. Some of these, especially Comprehensive Community Initiatives (CCIs), have given rise to the development of new evaluation approaches, responding to what Hollister notes (Brookings Institution, 1998), a mismatch between traditional evaluation designs and the expectations of rigor and utility that are placed on evaluations of complex interventions. CCIs can be best characterized as social interventions that employ a variety of strategies (mixing, for instance, direct service with efforts to change policymakers’ decision-making) and seek impacts on individuals, households, systems and communities (Kubisch, Weiss, Schorr, & Connell, 1995).

Some researchers have outlined the particular difficulties associated with evaluating CCIs and other complex social initiatives (Kubisch, Fullbright-Anderson, & Connell, 1998, p. 4). First, they note, an appropriate comparison group is elusive. Funders of CCIs purposely choose some communities over others precisely because they seem poised for change, or have strengths on which an intervention can be built. This selection strategy complicates the researchers’ task of finding appropriate comparison communities. Previous work suggests that matching communities on demographic and economic data often does not “hold-up” when other qualitative
The Case of the Urban Health Initiative

experiences are compared (Hollister & Hill, 1995, p. 2). Also, the creation of community comparison groups is hampered by the lack of recent data and time trends. Hollister and Hill (1995) note that these concerns apply to the selection of comparison groups for a wide range of evaluations, but CCIs complicate these problems further.

Second, as noted above, specifying the intervention to be evaluated can be difficult since such initiatives seek to ameliorate multiple social and economic problems and employ a variety of strategies. They are designed to respond to local conditions, and thus, may change their approaches midway through the intervention period, abandoning some strategies in favor of others, further complicating program definition.

In recent years, a variety of evaluation strategies arose that can be placed under the umbrella of Program Theory Evaluation (Rogers, Petrosino, Huebner, & Hacsi, 2000). As Rogers et al. have described, these strategies share two features: the identification of an explicit theory or model underlying the intervention and an evaluation design guided by the identified theoretical model. One such strategy, “theory of change,” has been promoted as a solution to the problems associated with evaluating CCIs, by specifying what program implementers believe will happen as a result of their intervention (Stufflebeam, 2001; Weiss, 1997). Theory, in this context, refers to the specific guiding model of those responsible for developing, directing, and implementing the program. In this approach, evaluators work with program staff to make explicit the theory and assumptions implicit in the programs they are undertaking. That is, using a theory of change approach, the evaluator seeks evidence that the assumed (or theorized) links between program activities or processes and the desired results have been borne out by experience. Researchers may then compare the expounded program theory about how an intervention will unfold to the observations evaluators make about how it does unfold.

Theory of change and other related approaches provide an important and promising framework for understanding CCIs, specifying the intervention, and helping to discern “lessons learned.” However, many argue that these approaches do not, in and of themselves, resolve the problem of the counterfactual (Hollister in Brookings Institution, 1998; Cook, 2000). Critics argue that knowing the program theory is not a substitute for empirically measuring program outcomes and contrasting them with a counterfactual condition (Peter Rossi in Brookings Institution, 1998). In fact, few program theory evaluations have tried to address the issue of attribution of causality (Rogers, 2000).

In debates about the relative merits of program theory evaluation, some have suggested that experimental and quasi-experimental designs are compatible with theory-based approaches. For example, Cook calls for the use of theory-based evaluation within randomized experiments (Cook, 2000). Indeed, in promoting new methods to address the problems associated with evaluating complex community initiatives, some have suggested enriched approaches that build on the strengths of both quasi-experimental and theory-based evaluation. Granger argues that theory of change approaches can be made more powerful by blending research designs and using multiple methods of inquiry (Granger, 1998). Such enhanced designs may meet Chen’s challenge that evaluation be responsive to the needs of stakeholders for usable information in shorter time frames, and produce credible results that allow for their generalizability and replicability (Chen, 1990). In considering such an integrative approach, Chen, Granger, and others note that the need for a counterfactual occurs throughout the implementation period, not only when assessing program outcomes. That is, complex social processes require evaluators to specify causal mechanisms and identify rival hypotheses throughout the intervention. This approach should facilitate program replication because sources of success or failure can be identified (Hacsi, 2000).
Large-scale empirical examples of program theory evaluation are rare, despite the extensive debate about its promise and its limitations (Rogers, 2000). Examples that try to blend experimental or quasi-experimental with theory-based approaches are, therefore, rarer still. In this case, we have constructed an enriched design, integrating the theory of change approach with a comparison group. This integrated approach offers two points of comparison by which underlying theoretical assumptions can be tested. First, the theory of change approach allows comparison of program experience to program expectation. Second, the comparison group design allows comparison of program experience to the experience of other like cities. Using theory of change embraces the complexity of the initiative, focuses more attention on links between specific processes and proximate outcomes, and encourages greater clarity in program design. Using a comparison group design strengthens our ability to assess the question of the counterfactual: would the desired activities and outcomes have occurred in the absence of UHI?

THE URBAN HEALTH INITIATIVE

In developing and launching their UHI in 1995, the RWJF recognized that although a number of promising approaches to improving the health and safety of children had emerged, most had succeeded either in improving outcomes in a single, categorical area of concern or in improving the health and safety climate of a single neighborhood. None had been sufficiently broad-based to significantly improve the health and safety of young people, measured against a number of indicators, across an entire city or metropolitan area. The UHI made a 10-year commitment to take on this challenge. The RWJF began an intensive and lengthy selection process to identify cities that, despite significant levels of distress, would have the necessary capacity and support to move forward on this effort; that is, they were “ready to make change.” The RWJF invited twenty cities initially to compete for funds by convening a single multi-sector planning group and identifying a lead agency to oversee the planning process. Eight of these groups were funded for a 2-year planning process that began in 1996. After this 2-year period, the groups in each city were expected to submit a proposal that described the city’s strategic plan to improve the health and safety problems of young people by focusing on public and private systems. Five city groups (i.e., Baltimore, Detroit, Oakland, Philadelphia and Richmond) were funded to go forward with implementing their plans for eight years beginning in 1998. At the end of the 10-year funding period, the Foundation expects measurable improvement in the health and safety outcomes for children in each of the cities as a result of the efforts described in these plans.

In many ways, the UHI resembled other CCIs, except that the geographic scale is larger. Building on the experiences and theories of CCIs, the RWJF called for a collaborative process for planning, implementing, and sustaining the initiative. Each of the citywide groups was allowed to identify for itself the issues that most threatened the health and safety of young people in their areas, and to come together to develop and implement new solutions to address these issues. As is typical with CCIs, the UHI plans were expected to be horizontally and vertically complex, and funds were intended to create systemic change, rather than to deliver services. The similarities between CCIs and the UHI led us, as evaluators, to use a theory of change approach.

THE EVALUATION OF UHI

Researchers at New York University were enlisted to evaluate this national initiative from the onset. The purposes of the national evaluation were to determine whether and how the UHI
sites were able to effect change on a range of health and safety outcomes for young people over the course of the initiative. The evaluation design is guided by three research questions:

- To what extent, and in what ways, can a foundation-sponsored initiative serve as a catalyst for a cross-sector collaborative process?
- To what extent, and in what ways, can a collaborative process result in meaningful changes in policies and programs designed to serve children and youth in urban settings?
- To what extent do these changes in policies and programs improve the health and safety of children and youth in urban settings?

As noted, we chose a theory of change approach to answer these research questions, and began our work by developing a theory of change with the funders of the intervention, that is, RWJF staff, to serve as a framework for our evaluation design. Of note, the theory applied to UHI as a whole, not to a single site’s efforts. As we worked to make explicit the underlying program theory, we were guided by a set of principles that had been espoused by the RWJF throughout the site-selection and planning periods. Although the groups were given considerable latitude in selecting problems, developing approaches, and designing their plans, the RWJF and its designated National Program Office (NPO) for the initiative described a set of core principles that they believed should guide the local efforts forward. These included:

- support and development of local leadership;
- use of a collaborative, multi-sector process for identifying critical issues and their solutions;
- incorporation of data-based decision-making;
- emphasis on changing public and private systems;
- crossing political boundaries with a regional approach;
- the development of a central communications campaign.

Building on these core principles, the national evaluation team developed a more fully explained theory of change for the UHI. An abridged version of this program theory appears in Figure 1.

Focusing on the earliest steps in the theory, one can see that the RWJF began with a set of underlying assumptions that build on a broad range of theories for political and organizational change. The RWJF, for example, anticipated that the use of a multi-sector planning process, with community involvement, would lead to a common understanding of the problems facing youth in that city. Further, the RWJF posited that the use of a “data-driven” planning process would lead to a greater ongoing use of analytic tools and best practice models. Using a theory of change approach, we could assess the progress of the initiative early on, by determining the degree to which these underlying theoretical assumptions, or links, were borne out in practice.

We determined which steps, or assumed links, along the program theory would be most important to measure, and chose methods to do so. The broad range of complex constructs that are found within UHI’s program theory necessitated the use of multiple methods of inquiry. These methods include:

- Annual visits to each of the five UHI program city sites, and to selected UHI comparison city sites;
- Annual key informant interviews with 6-10 civic leaders in each UHI and comparison city;
Figure 1. Abridged Theory of Change for The Robert Wood Johnson Foundation’s Urban Health Initiative.

were drawn in each UHI program city and its surrounding suburbs, and in the UHI comparison cities and their suburbs.

- Collection and analysis of annual administrative data from 1990 through the last year of project operations (2005) on a wide range of contextual and outcome indicators.

Some constructs are best measured by a particular method. For example, “a sense of urgency surrounding youth issues” is best measured through the telephone household survey,
while assessing “emerging leadership” requires key informant interviews. In other cases, two measures are used to provide a stronger, triangulated assessment of a single construct. For example, measuring “an improved allocation of resources” will require the use of both key informant interviews and the fiscal analysis.

The national evaluation team then sought to incorporate a traditional quasi-experimental design. Rather than using the RWJF’s theory of change as the only basis for comparison (e.g., did the sites engage in a multi-sector planning process and did this process result in a shared vision of youth problems?), the national evaluation team set about identifying a “like group” of comparison cities. We decided that we needed a comparison group that would allow us to rule out many of the rival hypotheses to explain observed changes, and one that would allow us to better address the counterfactual, in regard to both interim and final outcomes. For example, by adding a comparison group, we could test whether the multi-sector planning process was unique to UHI cities (and, therefore, attributable to the initiative). And, further, we would know whether a shared vision of youth problems was emerging in other cities, regardless of the presence of multi-sector planning.

USING CLUSTER ANALYSIS TO IDENTIFY A COMPARISON GROUP

In order to blend theory-driven and comparison group approaches, we believed it essential that some of the baseline features of the cities participating in UHI be integrated into the selection of the comparison cities. The small sample of UHI cities (5) necessitated the use of a matched comparison group; if the UHI and comparison cities had been chosen at random from the top 100 cities, there is a good chance that the program cities and comparison cities would differ considerably. For example, whereas the 20 invited cities had seen their population decline between 1990 and 1997, the rest of the nation’s cities, on average, had experienced growth. In the 20 invited cities, 27% of the households had incomes below $15,000 and 38% of the population was Black; in the rest of America’s large cities, only 21% of the households had such low incomes and only 20% of the population was Black.

Because random assignment was not possible, we needed an alternative method for identifying an appropriate comparison group. We sought to identify a group of cities that shared characteristics similar to the program cities, and employed cluster analysis to select these. Cluster analysis, first used in biology to categorize plants and animals according to like characteristics, is a multivariate procedure for detecting natural groupings in data. Clusters are based upon dissimilarities or distances between observations, calculated along single or multiple dimensions. Assuming the UHI program cities do represent some “like group,” cluster analysis allowed for the identification of a group of American cities similar to this UHI cluster. In this case, we made four guiding methodological decisions within the cluster analysis that were based on the underlying social and economic conditions of the UHI cities.

First, selection and weighting of analysis variables, as well as the decision to conduct analyses at two geo-political levels, had to reflect the program’s “target population,” that is, economically distressed U.S. cities. Moreover, the RWJF considered the underlying economic and social context within each city as important forces on youth well being. There is ample evidence in the literature that these variables are directly correlated with measures of health and safety, and the ability of cities to make effective change (Gruenewald, Treno, Taff, & Kitzmiller, 1997; Hauser, Brown, & Prosser, 1997; Lippman, Burns, & McArthur, 1996; U.S. Department of Housing and Urban Development, 2000). For example, a diminished tax base
and an increased concentration of people in poverty often accompany population loss. Therefore, cities that have experienced large reductions in population may experience an increased need to provide services while having diminished capacity to do so.

Second, we chose to examine the comparability of the UHI and comparison cities, in regard to their ability to create change, through qualitative interviews post hoc (we describe this below). Using cluster analysis constrained our use of qualitative information, and, therefore, did not allow us to match the comparison cities in regard to their “readiness for change.”

Third, we rejected the notion of using outcome measures for the cluster analysis. To include outcome measures would not permit tests of whether comparable baseline conditions give rise to poor outcomes or solutions. Perhaps most importantly, with comparison cities chosen on the basis of context, we can assess program effects on youth outcomes, above and beyond the underlying conditions.

Finally, we chose to treat the comparison cities as a group, rather than a series of five matched pairs. This was consistent with the RWJF’s primary concern with the success or failure of the initiative, rather than that of any single site. The RWJF’s approach focused on the UHI cities as a group, providing, for example, technical assistance to the group through national meetings and encouraging a great deal of sharing of information between the cities.

Further, by treating the UHI and comparison cities as two groups (while investigating individual differences), we alleviated some of the inevitable tension that arises between program and evaluation; our evaluation focus is on UHI, as a national initiative, rather than any particular site.

The UHI stresses broad-based approaches—spanning entire city or metropolitan areas and concentrating on multiple outcomes—in order to significantly improve the health and safety of young people. RWJF recognized that many past attempts to improve outcomes for youth had been tried only at the neighborhood level; they were interested in tackling outcomes for young people across cities, and large cities at that. To reflect the levels of distress and social characteristics of populations within these cities, we used six measures. Intercensal estimates of the following demographic characteristics for 1997 were used in selecting comparison cities: population, population change from 1990, percent of the population that is Black, population density, percent of household with income less than $15,000, and percent unemployed in 1990. As already noted, these characteristics in urban areas have been associated with high rates of youth related health and safety problems. To reflect the emphasis on “scale” in our selection process, absolute population size was weighted two times more than the remaining measures in our analyses.

The RWJF was also attentive to the importance of the relationship of cities to their suburbs in addressing the needs of youth. As indicated earlier, the UHI cities had suffered substantial loss of population, with much of this population migrating from the city out to its suburbs. Researchers investigating the interdependence of central cities and their suburbs have argued that regional approaches to solving social problems should be pursued (Rusk, 1996; Savitch & Vogel, 1996; Swanstrom, 2001); RWJF embraced these ideas in its thinking about UHI. Thus, using the comparison group to assess whether such efforts were feasible, underway, and finally contributing to improvements in youth outcomes, required the selection of cities with similar relationships to their suburbs. Therefore, we collected three measures of the central city in relation to its “suburban ring” for each observation in our analysis. Intercensal estimates of the suburban ring’s population in 1997, the change in that population 1990–1997, and the percent of that ring’s population that was Black in 1997 were included to understand the regional context.

We collected data for urban areas defined by metropolitan statistical areas (MSA) with a population greater than 500,000 and municipalities with a population greater than 180,000. The final sample consisted of 94 urban areas, including the five UHI cities. Using hierarchical clustering...
The Case of the Urban Health Initiative

Figure 2. Cities “Closest” to Philadelphia.

cluster analysis, we calculated the squared Euclidian distances (or proximity scores) from the z-scores of the variables. With the distance between each city in the sample and every other city, we assembled a group of cities as “close” as possible to the UHI cities. As an example, Figure 2 is a graphical presentation of the distances between Philadelphia and its 10 closest neighbors.

We conducted two rounds of analysis, clustering with solely central city measures and clustering with regional measures. We considered for inclusion in the comparison group those cities that were among the closest to each UHI city. Of note, there were several instances where the UHI cities clustered among themselves, thus reinforcing our observation that such variables were relevant to the UHI selection process and reinforcing our decision to focus on the UHI cities as a group, rather than matching pairs of individual cities. For the sake of manageability, the final comparison group was narrowed to 10 cities (see Table 1). Both geographic distribution and data availability were considered in the selection process. Washington, DC was included, although it was found comparable only when using central city measures; its role in the evaluation remains uncertain, as will be discussed later.

As a basic check on the successful application of cluster analysis, we compared our comparison group to the UHI cities on the demographic measures included in this analysis.

<table>
<thead>
<tr>
<th>TABLE 1. UHI Cities and Comparison Group</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>UHI Cities (5)</strong></td>
</tr>
<tr>
<td>Baltimore</td>
</tr>
<tr>
<td>Detroit</td>
</tr>
<tr>
<td>Oakland</td>
</tr>
<tr>
<td>Philadelphia</td>
</tr>
<tr>
<td>Richmond</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
TABLE 2.
Contextual Measures, Unweighted Mean Values

<table>
<thead>
<tr>
<th>Measures</th>
<th>UHI Cities</th>
<th>Comparison Group</th>
<th>Rest Urban Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population (municipality), 1997</td>
<td>730,748</td>
<td>395,711</td>
<td>557,346</td>
</tr>
<tr>
<td>Population density (municipality), 1997</td>
<td>7,149</td>
<td>6,612</td>
<td>4,242</td>
</tr>
<tr>
<td>Population change (municipality), 1990–1997</td>
<td>−5.8%</td>
<td>−5.6%</td>
<td>5.9%</td>
</tr>
<tr>
<td>Percent of Black (municipality), 1997</td>
<td>58.0%</td>
<td>43.7%</td>
<td>19.2%</td>
</tr>
<tr>
<td>Percent of households with income &lt; $15,000 (municipality)</td>
<td>28.7%</td>
<td>28.2%</td>
<td>20.7%</td>
</tr>
<tr>
<td>Percent of unemployed (municipality), 1990</td>
<td>10.9%</td>
<td>9.9%</td>
<td>7.1%</td>
</tr>
<tr>
<td>Suburban ring population, 1997</td>
<td>1,814,566</td>
<td>1,116,209</td>
<td>807,406</td>
</tr>
<tr>
<td>Change suburban ring population, 1990–1997</td>
<td>7.3%</td>
<td>1.59%</td>
<td>7.5%</td>
</tr>
<tr>
<td>Percent of Black ring, 1997</td>
<td>11.8%</td>
<td>12.3%</td>
<td>6.3%</td>
</tr>
</tbody>
</table>

Note. Data sources are described in National Evaluation of the Urban Health Initiative, 2000.

(see Table 2). As would be expected, the demographic and economic contexts for the UHI cities are similar to that of the comparison cities, and quite different from that of the rest of urban America. (For a more detailed analysis and description of the data sources see National Evaluation of the Urban Health Initiative, 2000.) One notable exception is that of population size. This is not a surprise because the cluster analysis was based on normalized measures. Although the absolute differences in population size appear large, the relative gap between UHI cities and the rest of urban America are much larger for our other variables. Therefore, despite the double weighting of population size, other variables had more influence in the clustering.

CHECKING THE VALIDITY OF THE COMPARISON GROUP

In response to the previously described criticisms made by Hollister and Hill (1995), we chose to check the validity of our selection in regard to a number of additional program-relevant measures. We used documents from, and our observations of, the RWJF selection process to provide some insight into their ideas of the “target group”—that is, the cities to which the findings of this evaluation might be generalizable. RWJF chose cities with demonstrable levels of distress among families and youth. At the same time, they wanted to give the initiative every chance to succeed, and sought cities uniquely "poised to make change." The presence of capable leaders, and demonstrated participation from across sectors was also deemed critical to "readiness for change." Ideally, then, a valid comparison group would include cities with similar problems that are similarly positioned to create positive change. That is, the salient characteristics of the “target group” should be integrated into the notion of comparability, and yet, as already noted, we could not capture many of these constructs within the cluster analysis.

We used data regarding health and safety indicators from the telephone household survey, as well as key informant interviews to explore the quality of the baseline match. While broad contextual measures of the cities’ socio-demographic conditions were used in the cluster analysis, we found that the conditions for youth, including health and safety outcomes, match far more closely for the initiative and comparison cities than for those in the rest of urban America.
Some exemplary variables are presented in Table 3. (For more details see National Evaluation of the Urban Health Initiative, 2000.) Results of the UHI 1997 nationally representative telephone survey of parents and youth, with over-samples of 750 parents and 750 youth in each of the five UHI cities, and an over-sample of 750 parents and 750 youth in the comparison cities, suggested the same comparability (see Table 4). Both adults and youth in the UHI cities expressed levels of concern about problems facing youth in these cities that are more similar to those in the comparison cities than to those in the 100 largest cities taken in their entirety. These administrative and survey data provided additional evidence that the types of problems in the UHI and comparison cities, and the depth of those problems, are similar. While the match is not perfect, clustering with socio-demographic data allowed us to identify a comparison group that strongly reflects this aspect of the underlying theory of the intervention.

Finally, we also looked to our key informants in the selected UHI and comparison cities to validate our comparison group. As part of our design, we had selected 6–10 civic leaders in each of the UHI and comparison cities to interview. These leaders typically included the city commissioners of health and police, the director of a citywide youth-serving non-profit organization, leaders in the local philanthropic community and business leaders with well-known civic interests, leaders in the school system, as well as other leaders suggested by these interviewees. Our first round of key informant interviews (conducted 1998–1999 at the end of the 2-year planning phase and the beginning of the implementation phase) asked leaders to identify major problems facing young people in their cities. Key informants in both the UHI

### Table 3

<table>
<thead>
<tr>
<th>Measures</th>
<th>UHI Cities</th>
<th>Comparison Group</th>
<th>Rest Urban Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent of female heads, 1990</td>
<td>37.2</td>
<td>33.9</td>
<td>22.1</td>
</tr>
<tr>
<td>Percent of high school graduates, 1990</td>
<td>65.9</td>
<td>69.4</td>
<td>76.1</td>
</tr>
<tr>
<td>Percent of poor kids, 1990</td>
<td>34.9</td>
<td>34.4</td>
<td>23.0</td>
</tr>
<tr>
<td>Violent crimes per 100,000, 1990</td>
<td>1,929</td>
<td>2,134</td>
<td>1,272</td>
</tr>
</tbody>
</table>

### Table 4

<table>
<thead>
<tr>
<th>Measures</th>
<th>UHI Cities</th>
<th>Comparison Group</th>
<th>Top 100 Cities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent of adult responses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local government performs fair/poor</td>
<td>67</td>
<td>63</td>
<td>50</td>
</tr>
<tr>
<td>City schools perform fair or poor</td>
<td>64</td>
<td>58</td>
<td>44</td>
</tr>
<tr>
<td>Believe unsupervised kids a big problem in their neighborhood</td>
<td>35</td>
<td>28</td>
<td>19</td>
</tr>
<tr>
<td>Percent of child responses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do not feel safe at school</td>
<td>19</td>
<td>17</td>
<td>12</td>
</tr>
<tr>
<td>Believe chances for teen becoming teen parent are high</td>
<td>43</td>
<td>45</td>
<td>32</td>
</tr>
</tbody>
</table>


Samples were weighted to reflect population proportions and to correct for varying numbers of residential voice lines.

Group of top 100 cities includes the UHI cities and the comparison group cities.
and comparison cities identified the same problems—substance abuse, crime and education were most commonly mentioned—and described similar collaborative efforts that pre-dated the Urban Health Initiative in both sets of cities.

In addition, key informants in these interviews were asked to assess the capacity for cross-sector involvement, and to describe other public and private efforts to address youth-related issues; these questions aimed to assess the cities’ “readiness to make change” for youth. Informants in the UHI and comparison group of cities noted many similar efforts to address youth issues. Similar levels of leadership, and similar types of leaders, were described by both UHI and comparison group informants. They also cited the same obstacles to developing effective initiatives to address young people’s problems: difficulty in getting and using data, the fragmentation of public and private systems, the duplication of services, and poor relationships with state governments. Based on our assessments of the key informants, we found no evidence to suggest that the UHI cities were more “ready for change” than the comparison cities, despite the fact that the RWJF believed it had selected cities on this basis. The findings of the interviews increased our confidence in the validity of the comparison group. They also, however, caused us to reconsider the inclusion of Washington, DC, because key informant interviews suggested that its unique political status could make change more difficult.

USING AN INTEGRATED APPROACH: PRELIMINARY FINDINGS

As already noted, evaluators need to consider the counterfactual, not only at baseline and with regard to the outcomes of an intervention, but also throughout the progress of the intervention itself. Efforts like the UHI have, as their premise, a vision of altering complicated social, political, and economic arrangements. Such arrangements are fluid and dynamic, and evaluators seeking to understand interventions aimed at altering these arrangements face very difficult problems of attribution.

We use our two points of comparison (program theory to program experience, and UHI to comparison cities) to test assumptions, or links, in the program theory. If the RWJF’s theory is accurate, and the model is effective, the support of RWJF and technical assistance from its NPO will encourage a multi-sector, data-driven planning process. This process would result in further multi-sector engagement, new civic leadership for youth issues, and the emergence of a common vision of youth problems and their solution. These elements represent early, essential steps in the RWJF’s theory.

Did the UHI cities engage in the planning process as expected? Did this process result in the kinds of engagement and vision that the program theory anticipated? Using site visits and key informant interviews, we worked with the sites to consider the assumed link between the planning process and its achievements.

We documented and assessed participation in the planning process through site visits, interviews, and archival review looking to see which sectors were included. We considered the strength of the participation and investigated the consistency and source of the leadership. We assessed the use of data in establishing the plans. While the planning process did not unfold exactly as anticipated, most of the key elements contained in the earliest stages of the theory (see Fig. 1) were found in each of the sites (see Silver, Weitzman, & Brecher, 2002, for a full discussion of the planning process and its effects). Using these same methods, we then looked to see whether or not the planning process was followed by sustained engagement by people from a range of sectors, by a growth in leadership on youth issues, and by the emergence of a
common vision of youth problems. The theoretical link, as anticipated by the theory of change, was borne out in practice.

Is there evidence, however, that these achievements were the result of the initiative? That is, was the planning process unique to the UHI cities? And do the UHI cities show greater engagement, leadership, and consistency of vision about their youth than the comparison cities? Key informants in the comparison cities described a range of current planning efforts to address youth issues. These efforts were focused on a single-sector (e.g., changes to social services case management), single neighborhood (i.e., broad in mission, but narrow in geographic scope), or single issue (e.g., citywide efforts to reduce youth homicide). Such activities contrast sharply with the kinds of citywide, multi-sector planning activities found in the UHI cities. Further, upon completion of the planning process, key informants in the UHI cities offered a more coherent presentation of youth problems, placed greater emphasis on youth issues, and were able to identify broad-based leadership for addressing them. The first stages and links, identified in the theory, were achieved and the comparison group design gives further evidence that the UHI contributed to their achievement.

This use of the double comparison, between theory and practice and between UHI and comparison cities, is integrated throughout the life of the evaluation. Although some aspects of the theory will not be subject to this double scrutiny, primarily because of limited resources, we have selected several critical assumptions in the program theory to test in this manner. Moving further along the theory (as presented in Fig. 1), we see that the RWJF posits that the increased levels of engagement and leadership, as well as the emerging common vision, will lead to a sense of both urgency and excitement, and a shared understanding of youth problems and their solution among both policy makers and the public at large. The theory of change approach encourages us to use site visits, key informant interviews, and the survey to document and assess the assumed link between engagement and leadership and broad public support for the selected youth agenda. But the use of the comparison group allows us to have greater confidence in our assessment. If the UHI is following along the path anticipated by its program theory, we should at this stage see increased concern for youth problems in the UHI cities and suburbs, as compared both to baseline and to the comparison cities. And we should likewise see an increased belief that government and the not-for-profit sectors are committed to addressing youth problems.

According to the program theory, at the completion of the 10-year initiative, we should see a measurable, positive impact on youth outcomes in the UHI cities, as well as a restored sense of optimism. With the addition of the comparison cities, we will be able to see whether those cities participating in UHI experience a greater degree of improvement. UHI and comparison cities began with a host of similar characteristics. If, over the life of this initiative, the experience in the UHI cities closely follows the path identified in the RWJF’s theory of change, and if the assumed links between the initiative’s activities and both proximate and long-term outcomes have been demonstrated, we will have evidence that the model works. If the path taken by the comparison cities differs from those of the UHI cities, and their final outcomes are weaker, we will have found enhanced evidence of a program effect.

DISCUSSION AND CONCLUSION

Evaluation research is at the mercy of the counterfactual. Even in a theory of change evaluation, the need to ask whether the documented changes would have occurred in the absence of the
program begs for comparison. At the same time, with initiatives as complex as the UHI, it is imperative to understand the expectation of program planners and implementers and to test their assumptions as part of the evaluation design. The two approaches must be integrated, so that evaluators can build the case that any impacts found may be logically ascribed to the program. And, rather than simply comparing final outcomes (as in a more traditional quasi-experimental design), the use of a comparison group in a theory of change evaluation must take note of links and milestones throughout the intervention.

In this evaluation, working under the assumption that the UHI cities represent a group with particular challenges and resources, we sought a method able to assemble a group of cities similarly positioned. After clustering our sample of urban areas using contextual variables, that is, by using demographic and economic measures for the city and its region, we effectively identified a 10-city comparison group. We are greatly encouraged by the consistent validity of this group when compared at baseline with the UHI group along a host of administrative and survey measures and expert opinions. In the earliest phases of this initiative, we see evidence that the initiative is unfolding as the theory would predict, and we see evidence that like cities are not undergoing a similar experience.

Our approach uses multiple methods of data collection and integrates features of a quasi-experimental, comparison group design with that of theory of change evaluation. Despite many remaining limitations, this integrated evaluation seems to provide greater confidence in assessing whether future changes in health and safety outcomes may have resulted from the Urban Health Initiative.

While our design will not allow us to definitively answer questions of attribution, it will allow us to provide reasonable explanations of observed changes, and insight into how change in economically distressed cities might be made to improve the lives of children and youth. Without both the theory of change approach and the comparison group, our reasonable explanations would be much poorer. With them, we hope to provide insight on not only whether change occurs, but also how it does.

NOTE

1. One hundred cities met the criteria for inclusion in the analysis sample; however, six cities were ultimately excluded from the cluster analysis due to data constraints: Chesapeake, Lexington-Fayette, Orlando, Plano, San Bernardino, and Scottsdale.

ACKNOWLEDGMENT

This work is funded by a grant from the Robert Wood Johnson Foundation.

REFERENCES


