Finding the Impact in a Messy Intervention: Using an Integrated Design to Evaluate a Comprehensive Citywide Health Initiative
Beth C. Weitzman, Tod Mijanovich, Diana Silver and Charles Brecher
American Journal of Evaluation 2009; 30; 495
DOI: 10.1177/1098214009347555

The online version of this article can be found at:
http://aje.sagepub.com/cgi/content/abstract/30/4/495

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
American Evaluation Association

Additional services and information for American Journal of Evaluation can be found at:

Email Alerts: http://aje.sagepub.com/cgi/alerts

Subscriptions: http://aje.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations http://aje.sagepub.com/cgi/content/refs/30/4/495
Finding the Impact in a Messy Intervention: Using an Integrated Design to Evaluate a Comprehensive Citywide Health Initiative

Beth C. Weitzman,¹ Tod Mijanovich,² Diana Silver,¹ and Charles Brecher²

Abstract
This article uses the evaluation of the Robert Wood Johnson Foundation’s (RWJF) Urban Health Initiative (UHI), a 10-year effort to improve health and safety outcomes in distressed cities, to demonstrate the strength of an evaluation design that integrates theory of change and quasi-experimental approaches, including the use of comparison cities. This paper focuses on the later stages of implementation and, especially, our methods for estimating program impacts. While the theory of change was used to make preliminary identification of intended outcomes, we used the sites’ plans and early implementation to refine this list and revisit our strategy for estimating impacts. Using our integrated design, differences between program and comparison cities are considered impacts only if they were predicted by program theory, local plans for action, and early implementation. We find small, measurable changes in areas of greatest programmatic effort. We discuss the importance of the integrated design in identifying impacts.

Keywords
theory of change, comprehensive community initiatives, quasi-experimental designs, child health, urban health

Introduction
The growing complexity of governmental and philanthropic efforts to address the health and social problems affecting poor communities challenges evaluators to identify better approaches for measuring program effectiveness. This is especially true for evaluations of efforts intended to leverage

¹ Steinhardt School of Culture, Education, and Human Development, New York University
² Robert F. Wagner Graduate School of Public Service, New York University

Corresponding Author:
Beth C. Weitzman, NYU Steinhardt, Pless Hall, 4th Floor, 82 Washington Square East, New York, NY 10003. Email: beth.weitzman@nyu.edu
changes in an array of policies and programs rather than to fund directly new, improved, or expanded services. Traditional evaluation methods were developed to evaluate traditional interventions that typically use a clearly specified model to address a single indicator and last only as long as the external funding. For many funders, such interventions seemed like a “band aid” for catastrophic outcomes that could be better addressed by changing underlying social conditions within a community; that is, by creating lasting changes in the places where families live and children are raised.

This broader focus on efforts to change public systems and policies falls under the umbrella of “Comprehensive Community Initiatives” (CCIs). Rather than directly funding new, improved, or expanded services, CCIs focus on the review and revision of programs and policies, strengthening of community institutions, and redirection of public and philanthropic dollars as the means to improve outcomes for families and children (Emshoff et al., 2007; Kubisch, Fulbright-Anderson, & Connell, 1998). The approach has also been popular in the United Kingdom, where it includes area-based initiatives (ABIs) aimed at reversing urban decline and health action zones (HAZ) aimed at improving health outcomes. These interventions pose the question: can philanthropic or public dollars be spent “upstream,” where policies and programs aimed at meeting the needs of poor urban families are formed, to catalyze community and institutional change and produce a “ripple” or “multiplier” effect, thus reaching many poor families on many indicators over a sustained period of time?

Despite widespread interest in and philanthropic support for CCIs, relatively few have been subject to rigorous evaluation of their impacts. The paucity of evidence regarding the impacts of CCIs may reflect the inherent difficulties faced by designing evaluations for interventions that focus on process and are nonprescriptive in their specific strategies or intended outcomes. How can traditional evaluation methods be extended or adapted to identify the impacts of interventions that are, by definition, big and messy? The evaluation communities in the United States and the United Kingdom have struggled with this question (Burton, Goodlad, & Croft, 2006; Chen, 1990; Davidson, 2000; Rogers, Petrosino, Huebner, & Hacsi, 2000). While some continue to argue for randomized approaches or traditional quasi-experimental designs with strong comparison groups (Bloom, Riccio, & Verma, 2005; Hollister & Hill; 1995; Rindskopf & Saxe, 1998), others have been critical of randomized designs. They have, instead, adopted a “theory of change” approach, defined as “a systematic and cumulative study of the links between activities, outcomes and contexts of the initiative” (Connell & Kubisch, 1998). Others reject the split between experimentation and theory-based designs (Cook, 2000), opting for designs based on quasi-experimental methods that use program theory to illuminate the “black box” of implementation (Judge & Bauld, 2001; Weitzman, Silver, & Dillman, 2002).

This article uses the evaluation of the Robert Wood Johnson Foundation’s (RWJF) Urban Health Initiative (UHI), a 10-year effort to improve health and safety outcomes in distressed cities, to demonstrate the strength of an evaluation design that integrates theory of change and quasi-experimental approaches, including the use of comparison cities. As we will discuss, this integrated design allows evaluators to identify more accurately potential program effects for subsequent impact analysis, thus strengthening the counterfactual argument beyond what would otherwise have been possible. The evaluation design was purposely kept flexible in the early years of the initiative; this allowed the sites’ perspectives and experiences to be reflected in the final analyses. (We use the word site to refer to the lead agency and other local governing structures that oversaw the work of UHI, and city to refer to the geopolitical entity.)

Previously published work indicated that the earliest stages of this initiative unfolded in a manner largely consistent with the program theory, and that the experience of the UHI sites during the planning and early implementation stages was different from that of the comparison cities (Silver, Weitzman, & Brecher, 2002; Weitzman et al., 2002). This article extends that prior work, focusing
on the later stages of implementation, and, especially, program impacts. We begin with a brief overview of the program and its underlying theory of change. We then review our strategic approach to its evaluation and methods for gathering data. Next, we discuss how the sites’ plans for action and their early implementation led us to refine the list of intended outcomes for the initiative and our strategy for estimating impacts. The fourth section summarizes our findings regarding the medium- and long-term impacts of the initiative and discusses how our integrated design allowed us to better estimate them. We end with a discussion of the strengths and weaknesses of this evaluation approach and the conditions under which it may be most beneficial.

Overview of the Program and Theory of Change

In the early 1990’s, RWJF, dissatisfied with its prior performance in creating widespread and enduring improvements in health outcomes within America’s largest and poorest urban areas, challenged itself to consider new ideas and approaches. In 1995, as a result of this challenge, RWJF launched UHI (RWJF, 1995). It had six distinctive features:

1. a focus on large, distressed cities, with deeply entrenched problems;
2. an emphasis on changing systems, with grant funding intended as “venture capital” to change public systems to produce better outcomes;
3. local choice regarding areas and strategies of focus;
4. the engagement of strong local leadership in multiple sectors;
5. a 10-year commitment, with funding of over U.S.$80 million for site activities, technical assistance, and program evaluation; and
6. a goal of measurable change, citywide, in multiple domains relevant to the health and safety of children and youth.

Initially, sites in eight cities were funded for 2 years to engage in citywide, multisector, data-driven planning (Silver et al., 2002). The foundation selected five sites—Baltimore, Detroit, Oakland, Philadelphia, and Richmond—to receive approximately U.S.$1 million per year to implement their plans, beginning in January 1997. While the Foundation stated that the change needed to be measurable and citywide, the amount of change that defined success was not specified.

RWJF’s theory of change for UHI embraced a political theory in which a multisector, data-driven planning process, engaging both the civic elite and the public at large and emphasizing a close reading of the literature on best practice models, would lead to a shared understanding of the problems facing youth and a common vision regarding the strategies to ameliorate them (see Weitzman et al., 2002, for more information about the theory of change and its development). Consistent with this belief in the importance of broad engagement, RWJF emphasized that UHI did not seek to affect the lives of only low-income people. Such an initiative, they argued, could not obtain the political support needed to “move the needle” citywide. Emphasizing the city as the focus of intervention represented a departure from the community focus more typical of CCIs (see Silver & Weitzman, 2009, regarding this difference).

The Foundation’s staff theorized that the process of citywide engagement would result in a sense of urgency and the political will to restructure policies and programs, that is, change systems, to better achieve their shared vision. Expenditures would be shifted from costly and ineffectual remedial or “corrective” approaches to more effective and lower cost “preventive” approaches. By engaging the public at large, norms and behaviors among parents and youth would be reshaped to reinforce and enhance these improvements in policy and programs. The Foundation believed that well-implemented upstream interventions would improve outcomes for children, citywide and in multiple domains, helping to restore optimism among families in these cities.
To support and integrate the local efforts in putting the theory into practice, RWJF created a national program office (NPO) to provide ongoing and intensive technical assistance to the sites and to monitor their activities and spending. RWJF funded the evaluation team to provide the sites with technical assistance to meet their local research and evaluation needs, as well as to conduct the national evaluation. Finally, midway through the initiative, RWJF funded a seminar series at Harvard University, to facilitate a conversation among UHI leadership at the local and national levels and prominent researchers in the areas of urban poverty and child well-being.

The Evaluation Strategy and Methods

The national evaluation team sought to address the challenges posed in evaluating a long-term, citywide, comprehensive initiative by combining a quasi-experimental design, including a group of comparison cities, with a theory of change approach (see Weitzman et al., 2002). While the comparison cities were not studied in equal depth as the UHI cities, efforts were made to understand their local landscape in regard to youth problems and their solutions. Furthermore, the evaluation, reflecting RWJF’s own emphasis, stressed looking at the initiative as a whole rather than deeply drilling down into any one site; this kind of site-specific evaluation activity was left to the sites and their local evaluators.

The theory of change, developed by the evaluation team through an iterative process with senior staff from RWJF and the NPO, was used to define the initial set of intended impacts. Given the length of the initiative, it was important that outcomes for the short, medium, and long term be delineated (see Table 1). The short-term outcomes focused on the process of engagement in the planning phase. Medium-term outcomes were best summarized by the initiative’s tag line, “working smarter for kids.” These outcomes included an enhanced sense of urgency regarding problems facing the cities’ children and creation of a shared vision about their solution. The long-term outcomes anticipated were improvements in life conditions for children and youth, improvements in a broad array of health and safety outcomes citywide, and restored optimism among families.

| Table 1. Outcomes Predicted by the National Theory of Change for the Urban Health Initiative |
|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|
| **Short-term outcomes: planning phase**        | **Medium-term outcomes: “working smarter for kids”** | **Long-term outcomes: improvement in health and safety outcomes** |
| Increased engagement across multiple sectors  | Enhanced sense of urgency regarding problems facing the city’s children | Improvements in life conditions for children and youth |
| Increased engagement among community participants | Creation of a shared vision for addressing youth problems | Improvements in a broad array of health and safety outcomes citywide |
| Increased engagement among civic elite         | Emergence of political will to make needed changes | Restored optimism for families |
| Increased use of data and best practice models for defining and addressing problems facing children and youth | Strengthening of policies and programs relevant to children and youth through the use of data, best practice, and reorientation of funding | Reshaping the norms and behaviors of young people and their parents |

To support and integrate the local efforts in putting the theory into practice, RWJF created a national program office (NPO) to provide ongoing and intensive technical assistance to the sites and to monitor their activities and spending. RWJF funded the evaluation team to provide the sites with technical assistance to meet their local research and evaluation needs, as well as to conduct the national evaluation. Finally, midway through the initiative, RWJF funded a seminar series at Harvard University, to facilitate a conversation among UHI leadership at the local and national levels and prominent researchers in the areas of urban poverty and child well-being.
In our integrated evaluation design, qualitative and quantitative data were used to test for program impacts. Qualitative implementation research examined relevant activities and processes to see whether the program unfolded as predicted by theory and differently from that in similarly distressed comparison cities; that is, whether medium-term outcomes had been achieved. Quantitative data, including survey responses from parents and youth and administrative data regarding citywide child health and safety outcomes, were used to test for medium- and long-term impacts. We employed a difference-in-differences approach to examine whether the UHI cities experienced greater changes in intended outcomes than did the comparison cities. Differences between these changes are, however, attributed to the initiative only if these outcomes were predicted by both program theory and implementation to be intended impacts; our reasons for this critical decision are discussed below.

We analyzed data from multiple sources, including:

- annual 2–3 day site visits to each city during which local theories of change were explored;
- participation in all national meetings for the initiative;
- interviews with 10–15 key informants in each city every 12–18 months;
- three waves of a random-digit dialed telephone household survey of adults and youth;
- analyses of public expenditures by the city, state, and federal governments to determine changes in the amount and purpose of monies spent on services to children (see Brecher, Searcy, Silver & Weitzman, 2004, for a description of these methods); and
- changes in multiple health and safety indicators over a 10-year period.

The Survey of Adults and Youth (SAY) was fielded in the five program and nine comparison cities in the 1998–1999, 2001–2002, and 2004–2005 school years. It asked parents and children a variety of questions taken from the National Longitudinal Survey of Youth, the Youth Risk Behavior Survey, Monitoring the Future, the National Health Interview Survey, and others, which focused on conditions and quality of life in respondents’ neighborhoods and cities; performance of city institutions; problems affecting youth; parent-child activities, relationships, and rule setting; school climate and activities; after-school activities; and other child-related issues.

SAY's sampling design generated approximately 1,500 adult interviews in each UHI city and 1,500 interviews in the group of comparison cities. Two thirds of the SAY sample were parents of at least one child younger than 19 years in the household, while the remaining one-third were in households without minor children. Up to eight callbacks were made to each household. SAY was offered in English and Spanish (and in Oakland, in Chinese). Among contacted households across all three survey waves, 89% of adults agreed to participate, and among participating parents, 74% agreed to have their child interviewed; virtually all these children agreed to participate.

The UHI evaluation also collected administrative data for the period 1998–2005 from various sources, including FBI Uniform Crime Reports (UCR), vital statistics (birth and death), and sexually transmitted infection (STI) data from the Centers for Disease Control (CDC), census and Current Population Survey (CPS) data, and educational data from the National Center for Educational Statistics (NCES). Raw event counts were tabulated from administrative sources and converted to citywide rates for various outcomes using denominators interpolated or extrapolated from census or American Community Survey (ACS) data.

Incorporating the Sites’ Plans and Early Implementation into the Evaluation Plan

After the local plans were developed and implementation had begun, we revisited our original list of outcomes and our analytic strategy. This flexibility was critical for capturing the significant changes
that occurred to the initiative as the Foundation’s theory was operationalized in practice. This pro-
cess of revision was necessitated by the nonprescriptive and long-term characteristics of the initia-
tive itself. The Foundation anticipated that the sites would move forward based on local conditions
and opportunities; our evaluation plan needed to be sufficiently agile to adjust to these conditions.

Refining the Outcomes

While the theory of change was used to make a preliminary identification of the initiative’s intended
outcomes, we used the sites’ plans to refine the list and to help identify relevant measures for them;
this was critical, given the nonprescriptive nature of the initiative in regard to policy targets and
actual plans of action. Findings from early implementation research were further used to confirm
or revise the medium- and long-term outcomes, such that they reflected the focus of actual program-

matic activity.

The sites were funded for implementation only after the planning process was extended; during
this time, they were required by RWJF and the NPO to reconsider, and narrow, their originally iden-
tified goals and objectives and to estimate the number of young people to be reached by their pro-
posed strategies. These requirements represented a significant departure from the original theory of
change, with its emphasis on politics and broad systems change. As a result of these new require-
ments, the timetable was changed, the program became narrower, and the emphasis on community
norms was dropped.

The final plans identified those youth problems that were to be the focus of the local efforts and
the specific strategies to address them. All five sites chose to implement the model for change spe-
cifically in regard to after-school expansion and four also chose to focus on youth violence reduc-
tion. Infant health improvement through early family intervention and increases in educational
achievement were also the focus in two of the plans. Although the sites chose many of the same areas
of focus, it is noteworthy that they did not adopt identical strategies. Questions were thus added to
the later waves of the household survey to reflect the emphasis on out-of-school time demonstrated
in the plans and early implementation. Table 2 presents the approaches and the principal actions pro-
posed in each site to meet those goals.

Differences in the local plans for action were the result of differences in local conditions and
political culture and also the lack of consensus among experts about “best practices” for improve-
ments in these areas (Silver et al., 2002). Despite these differences, each strategy reflected multiple
elements of the medium-term outcomes identified in the theory of change, including the emphasis on
best practices, increased sense of urgency and political will, and strengthening of youth-serving
organizations.

Overall, the areas chosen for intervention were not unique among these distressed cities; indeed,
they reflected areas targeted by federal agencies, large and influential foundations, and nationally
recognized nonprofits (Silver et al., 2002). Rather, what made the UHI sites distinctive was the pro-
cess used to select priorities and strategies. Thus, the integrated design with its comparison cities
remained a strong counterfactual test of whether UHI’s approach to these youth problems would
result in changes in outcomes.

Table 3 summarizes the indicators we identified as evidence of the achievement of the major out-
comes for each stage of the initiative, given the sites’ plans for action. While only qualitative evi-
dence was used to assess the extent to which those indicators associated with the planning phase had
been achieved, both qualitative and quantitative indicators were identified to assess whether the
anticipated medium-term outcomes had been achieved. The long-term outcomes were assessed
solely through quantitative indicators. However, our interpretation of the findings on long-term out-
comes benefited from the qualitative evidence regarding medium-term outcomes; the “black box”
linking inputs to outcomes was no longer empty.
<table>
<thead>
<tr>
<th>Site</th>
<th>Year chosen</th>
<th>Target group</th>
<th>Approach</th>
<th>Principal actions taken</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baltimore</td>
<td>1997</td>
<td>K-12 students</td>
<td>Expand after-school participation</td>
<td>Created quality standards and provided technical assistance to meet them; lobbied for additional public and private dollars for new slots; worked with partners to identify and implement best practices; evaluated funded programs, and linked funding to program performance.</td>
</tr>
<tr>
<td>Detroit</td>
<td>1999</td>
<td>K-12 students</td>
<td>Expand after-school participation</td>
<td>Created Web site and other venues for parent information; publicized benefits and availability of after-school; worked with public partners to streamline access to public facilities; created citywide database of programs and enrollment; facilitated TA to new and emerging programs.</td>
</tr>
<tr>
<td>Oakland</td>
<td>2001</td>
<td>Middle school students</td>
<td>Improve awareness of after-school among parents and strengthen capacity for management and accountability</td>
<td>Blended behavioral health dollars with other funding streams; placed coordinators in targeted middle schools; facilitated linkages between schools and providers.</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>2001</td>
<td>K-12 students</td>
<td>Link at-risk middle school students to appropriate services and after-school programs and inform parents of all students of existing school-based opportunities</td>
<td>Established quasi-governmental intermediary organization to create new funds, eliminating existing obstacles and providing TA.</td>
</tr>
<tr>
<td>Richmond</td>
<td>1998 (de-emphasized in 2003)</td>
<td>Elementary school students</td>
<td>Integrate reading support activities into existing after-school programs in a variety of settings</td>
<td>Worked with public and not-for-profit partners to provide reading support activities through program-specific curricula.</td>
</tr>
<tr>
<td>Site</td>
<td>Year chosen</td>
<td>Target group</td>
<td>Approach</td>
<td>Principal actions taken</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Baltimore</td>
<td>2001; young adult strategy never fully implemented</td>
<td>Youth ages 13–17 years; companion strategy targeting adults 18–24 years of age</td>
<td>Reduce youth homicide and juvenile arrests through community-based approach targeting high-risk youth and providing coordinated-case management</td>
<td>Studied and identified best practice models; linked police, health department, and experts; raised private funds to seed efforts</td>
</tr>
<tr>
<td>Detroit</td>
<td>1999, dropped in 2001</td>
<td>Targeted communities with higher crime rates</td>
<td>Increase citizen involvement in community safety activities</td>
<td>Assisted police department and other public agencies with community activities</td>
</tr>
<tr>
<td>Oakland</td>
<td>Varied until 2001</td>
<td>Three Groups: first time offenders, ages 10–17 years; middle school students grades 6–8; young children who had witnessed violence at home or in the communities</td>
<td>Use mental health and related services to reduce youth crime and victimization, and its sequelae. Provide social, psychological and school services to first time offenders. Implement conflict resolution curriculum in targeted middle schools. Link families where violence has occurred in the presence of very young children to counseling services</td>
<td>Identified best practices, convened key stakeholders, raised funds, provided TA, conducted evaluations of efforts</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>2001</td>
<td>Youth in targeted, high crime neighborhoods</td>
<td>Reduce youth homicide and violence through improved policing strategies in targeted districts</td>
<td>Studied and identified best practice models, facilitated cross-agency planning, tracked relevant data, and evaluated impact</td>
</tr>
</tbody>
</table>
Table 2  (continued)

<table>
<thead>
<tr>
<th>Site</th>
<th>Year chosen</th>
<th>Target group</th>
<th>Approach</th>
<th>Principal actions taken</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baltimore</td>
<td>1998</td>
<td>Pregnant women and families with new babies in high-risk neighborhoods</td>
<td>Reduce infant mortality through home visitor models</td>
<td>Raised funds, identified best practices, evaluated pilot programs, monitored performance</td>
</tr>
<tr>
<td>Richmond</td>
<td>1998</td>
<td>Babies and very young children in at risk environments</td>
<td>Reduce infant mortality and accidental injuries through: home visitor model and United Way’s Success by Six campaign; improvements in quality of child care settings.</td>
<td>Identified best practices; convened key stakeholders; provided staff development and TA</td>
</tr>
</tbody>
</table>

Improve educational outcomes, especially early reading

<table>
<thead>
<tr>
<th>Site</th>
<th>Year chosen</th>
<th>Target group</th>
<th>Approach</th>
<th>Principal actions taken</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baltimore</td>
<td>1998, never fully implemented</td>
<td>Students, K-3</td>
<td>Increase the proportion of kids reading at grade level by grade 9 through improvements in instruction within elementary schools</td>
<td>Identified best practices; assembled data; enlisted stakeholders</td>
</tr>
<tr>
<td>Richmond</td>
<td>1998, most elements de-emphasized by 2003</td>
<td>Elementary school students</td>
<td>Increase the proportion of kids reading at grade level by grade 9 through: tutoring program for at-risk readers; adoption of citywide reading curriculum; read-to-your child promotional campaign</td>
<td>Identified best practices; convened key stakeholders; provided TA; launched public service campaign and book collection and distribution</td>
</tr>
</tbody>
</table>
Refining Our Analytic Strategy

As previously described, our analytic strategy rested on a difference-in-differences approach, looking at changes over time in the UHI cities relative to the comparison cities in indicators from the SAY and administrative data. After reviewing the local plans and reflecting on the planning process and earliest phases of implementation, we adjusted our strategy for analyzing the survey data, while leaving our approach to the administrative data intact.

**Analysis of survey data.** Using logistic regression, program impacts were estimated as odds ratio (OR) coefficients on the interactions between survey wave and UHI/comparison city indicators. The impact is the change in the probability of a given response relative to the trend in the comparison cities. Because survey data were collected at the individual level, we chose to analyze them at the individual level with appropriate adjustments for the clustered nature of the data. While this framework was left unchanged, five unanticipated features of the plans and early implementation led to adjustments in our evaluation strategy.

First, the narrowed and site-specific nature of the plans suggested that impacts would most likely be found in the sites’ chosen areas of focus (e.g., after-school activities) rather than across the board as predicted by the original theory of change. We considered a statistically significant difference in differences to be a program impact only if (a) it was predicted by the causal theory of the program, (b) it occurred in an area theorized to be affected by program activities that we were able to document as having been at least partially implemented, and (c) it was part of a pattern of, and did not contradict, other impact findings. The second criterion reflected our response to the narrowing of programmatic focus.

We estimated several different impact models within each domain. We employed data reduction techniques within each response domain and estimated impacts on two resulting variables. One variable was the first extracted factor from a principal component analysis of the variable domain (Jolliffe, 2002). The second variable was the average of z-transformed variables in the domain (see, e.g., Kling, Liebman, & Katz, 2007). Impact findings using these summary variables were consistent with impact estimates using the raw variables. For ease of interpretation, only the impact estimates based on untransformed measures are presented. We also estimated alternative models, including models with robust standard errors adjusted for within-city dependence of observations and hierarchical linear models with city-level random effects. Coefficients and p values were very similar using these alternative methods and are not shown.

The high bar set by the three criteria enumerated above seemed especially important once we recognized the complex nature of the sites’ plans and early efforts. Given the presence of other youth-related efforts in both the UHI and the comparison cities, we wanted to credit the initiative with improvements in youth outcomes if and only if the preponderance of evidence pointed in that direction. This represented a trade-off between two types of error: Type 1, where we would have falsely credited UHI with improvements it did not create by setting the bar too low; and Type 2, where we would miss crediting UHI with improvements it did create by setting the bar too high. The risk of Type 2 error may be particularly great if one believes that spillover effects to outcomes outside the areas of focus were especially likely. We chose to lower the risk of Type 1 error while increasing the risk of Type 2 error because we believed that widespread skepticism about the benefit of CCIs demanded this approach and because our close read of UHI’s theory of change, and UHI’s implementation, suggested that originally anticipated spillover effects would not likely occur in the period of time under study.

Second, given the initiative’s nonprescriptive philosophy, the sites were given latitude in the timing of planning and implementation activities (Silver et al., 2002). This meant that in many programmatic areas, implementation did not begin until 2001 or later; such variation in timing was not
Table 3. Outcome Indicators for the UHI Evaluation

<table>
<thead>
<tr>
<th>Type of evidence</th>
<th>Short-term: planning</th>
<th>Medium term: working smarter</th>
<th>Long-term: improvements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qualitative:</td>
<td>Documentation that planning process engaged multiple sectors, civic elite, and community residents, and relied on data and best practices for decision making&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Documentation that implementation of UHI plans included activities such as town hall meetings and mayoral forums to create urgency, a common vision, and political will. Comparison with other distressed cities using key informants’ assessments of the extent of common vision, leadership, and political will for addressing youth problems. Documentation that UHI sites undertook activities such as city performance metrics and evaluations of new programs that used data and best practice models to strengthen policies and programs. Fiscal analysis to see whether public expenditures for youth increased and moved toward more preventive strategies</td>
<td>None</td>
</tr>
<tr>
<td>Key informant interviews</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site visits</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Media scan</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public expenditure analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Quantitative: | None | Comparison with other distressed cities in regard to changes in youth homicide, infant mortality, accidental injury, educational achievement, and so on. Comparisons with other distressed cities of survey responses in regard to changes in parents’ and youth’s assessments of neighborhood conditions; out-of-school time services; youth safety; school conditions; conditions for young children | |
| Survey | | | |
| Vital statistics data | | | |
| Education data | | | |

<sup>a</sup> See Silver, Weitzman, & Brecher, 2002 and Weitzman, Silver, & Dillman, 2002 for a discussion of the evidence in this regard.
anticipated at the onset of the initiative. This required us to rethink the notion of a firm baseline and enabled us to use the 2001–2 SAY survey wave as baseline data for selected outcomes, especially in regard to questions concerning out-of-school time. In the analyses of the SAY data, the baseline year was the earliest school year for which we collected relevant survey items: either 1998–9 or 2001–2.

Third, the extent of demographic changes that were to occur in these cities over the life of the initiative was not appreciated at its start. It was hoped that UHI would influence the demographic mix by restoring optimism among families, who might then choose to remain in these shrinking cities. Quickly it became clear that larger social and economic forces, most especially immigration, regional housing dynamics, and job markets, would swamp any impact that UHI might have on a city’s demographic mix. We realized that our statistical models would need to include demographic characteristics for the respondents and their neighborhoods, to adjust for sampling variations across survey waves and, more importantly, to “net out” demographic changes in the cities over the 10 years, because these changes could not reasonably be attributed to the initiative itself. Thus, we included in our models both individual-level covariates (White race; high school degree only, and college degree; household income >50k/year; marital status; and employment status) and zip-code level covariates (Census 2000 percentage of the zip’s residents who were under 200% of poverty, who had a college education, who were non-Hispanic black, who were 5 years old or younger, and who were homeowners; and the percentage of houses in the zip code that were vacant).

Fourth, given the variation across cities in focus and strategies, we needed to reconsider our decision to look for impacts at the site level. While we continued to focus on the impact of the initiative as a whole, we also decided to estimate impacts for individual cities to test whether impacts occurred in site-specific areas of intervention. Individual site outcomes were compared to average outcomes in the comparison cities for these estimates.

Finally, although UHI was intended to engage people across the economic spectrum and its interventions were not targeted only to disadvantaged populations, the local plans for action and early implementation findings indicated that the sites sought to make their greatest impact within disadvantaged communities where youth problems were most concentrated. This shift in the theory needed to be incorporated into our analytic strategy. We decided to conduct impact analyses for subgroups based on the educational attainment of the parent respondent; we chose not to use income because of the extent of missing data. While the original theory predicted changes in youth outcomes citywide, the plans and experiences within the sites predicted changes specifically among youth from lower SES families.

Analysis of administrative data. Despite these important changes in our strategy for analyzing the SAY data, we decided to make few changes in our strategy for analyzing the administrative data. The administrative data provide the broadest and most objective indication of whether health and safety outcomes improved in the UHI cities relative to other distressed cities over the life of the initiative. In theory, it is possible that improvements could be realized in the areas of greatest programmatic focus, but at the expense of other areas of need. Using a variety of indicators from the administrative data allows us to examine that possibility. Furthermore, from the standpoint of policymakers, these rates, unadjusted for population shifts, represent the most readily understood indicators of change; local leaders are interested in seeing improvements regardless of their source. We decided to stick with our original plan; impact estimates using administrative data were estimated as differences in differences, comparing the average change between baseline and follow-up years in UHI cities with average change over the same time period in comparison cities. For all administrative data measures, calendar year 1998 was considered the baseline, without adjustment for the slow roll-out of the program. Differences in differences in administrative data measures were not tested.
for statistical significance, because these measures were calculated using the full city populations rather than samples.

Our administrative data analysis thus complemented our analysis of SAY survey data. Whereas the latter was designed to detect narrowly defined outcomes in areas of programmatic focus, the administrative data analysis was intended to detect changes in overall child health and well-being.

What Did We Find?

Overall, the local implementation of UHI was roughly consistent with the program’s theory, even as it was slower and less “orderly,” that is, less linear than originally anticipated. In contrast with the comparison cities, whose leadership shared many of the same goals, the UHI cities benefited from a convening body whose primary function was not to provide services but rather to facilitate a coordinated effort to achieve shared goals. As a result, many of the intended medium-term outcomes were at least partially achieved.

Despite many obstacles, the UHI sites engaged in a variety of strategies to raise public perceptions and expectations, create a common vision, increase political will, and improve the accountability of city government to the public. Philadelphia published “report cards” listing indicators of child well-being for the city at large. Baltimore developed a “children’s budget” that analyzed public spending on services for young people. In addition, several sites organized candidate forums on children’s issues and voter registration campaigns to inform the public about referenda regarding youth-related policies. Senior staff and board members from each of the UHI sites also met regularly with local and state government elected officials, as well as those actively seeking office, to share the local UHI vision and plans for action.

UHI sites also engaged in formal and informal efforts to encourage a more data-driven, evidenced-based approach to making policy and selecting programs for funding and implementation. Each site reviewed research literature and engaged local and national experts to identify and promote “best practice” (rather than “most popular”) models. The sites worked extensively with city agencies and nonprofit groups to regularly collect, analyze, and share data related to population needs and program administration. They also attempted to gain a better understanding of where and how public dollars were being spent on services, with the expectation that this would allow them to realign funding.

All these UHI site activities were in dramatic contrast to the state of youth policy and program development in the comparison cities. Although the comparison cities were attempting to improve many of the same health and safety problems as the UHI cities, their efforts lacked coordination and consistency. As one key informant put it, there was no “systematic approach” in his city to youth-related issues and no “big umbrella group” attempting to coordinate disparate strategies. Civic leaders in the comparison cities talked about the absence of a champion for “data-driven” and evidence-based decision making to define and understand problems, select strategies, and monitor their outcomes, especially across city agencies and nonprofits. This contrast between the UHI and comparison cities indicates that, indeed, UHI sites found ways to “work smarter for kids” that would not have been pursued in the absence of UHI.

Our key informant interviews and survey results confirm that both civic leaders and community members responded to these efforts (see Table 4 for an overview of impacts on selected key indicators). The UHI sites presided over the emergence of a shared vision and increased civic engagement, and parents perceived that something different was happening in their cities. Relative to parents in the comparison cities, parents in the UHI cities became more engaged in civic matters (as measured by survey self-report of voting in local elections: OR = 1.45, p < .01) and more positive about their city’s efforts to improve youth well-being (OR = 1.24, p < .1). Parents in UHI cities
Table 4. Urban Health Initiative Impacts on SAY Survey Outcomes

<table>
<thead>
<tr>
<th>Domain and outcome</th>
<th>Odds ratio</th>
<th>Baseline wave</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Assessment of youth problems</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young people skipping school is a big problem in this city</td>
<td>1.48***</td>
<td>2</td>
</tr>
<tr>
<td>Dropping out of high school is a big problem in this city</td>
<td>1.22*</td>
<td>1</td>
</tr>
<tr>
<td>Teen pregnancy is a big problem in this city</td>
<td>0.82*</td>
<td>1</td>
</tr>
<tr>
<td>Young people vandalizing is a big problem in this city</td>
<td>0.74***</td>
<td>2</td>
</tr>
<tr>
<td>Problems will get better in next 5 years</td>
<td>1.14</td>
<td>1</td>
</tr>
<tr>
<td><strong>Civic engagement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Worked with others on community problems in past 12 months</td>
<td>1.07</td>
<td>1</td>
</tr>
<tr>
<td>Always vote in local elections</td>
<td>1.45***</td>
<td>1</td>
</tr>
<tr>
<td><strong>City and institutional responsiveness to kids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>City good/excellent for kids to grow up in</td>
<td>1.17</td>
<td>1</td>
</tr>
<tr>
<td>City effective at improving youth well-being</td>
<td>1.24*</td>
<td>2</td>
</tr>
<tr>
<td>Local government good/excellent</td>
<td>0.97</td>
<td>1</td>
</tr>
<tr>
<td>Local community organizations good/excellent</td>
<td>1.03</td>
<td>1</td>
</tr>
<tr>
<td><strong>Neighborhood conditions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neighborhood good/excellent place to raise kids</td>
<td>1.24*</td>
<td>1</td>
</tr>
<tr>
<td>Youth says neighborhood good/excellent place to raise kids</td>
<td>0.97</td>
<td>1</td>
</tr>
<tr>
<td><strong>Out-of-school time</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive about after-school programs</td>
<td>1.30*</td>
<td>2</td>
</tr>
<tr>
<td>After school availability good/excellent</td>
<td>1.42***</td>
<td>2</td>
</tr>
<tr>
<td>City effective at developing after-school</td>
<td>1.46***</td>
<td>2</td>
</tr>
<tr>
<td>Youth organizations good/excellent</td>
<td>1.07</td>
<td>1</td>
</tr>
<tr>
<td>Parks and recreation centers good/excellent</td>
<td>1.21*</td>
<td>1</td>
</tr>
<tr>
<td><strong>Safety and violence</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neighborhood has safe places for kids</td>
<td>1.24*</td>
<td>1</td>
</tr>
<tr>
<td>Youth agrees/strongly agrees she or he felt safe everywhere yesterday</td>
<td>0.92</td>
<td>1</td>
</tr>
<tr>
<td>Crime and violence a big problem in neighborhood</td>
<td>0.99</td>
<td>1</td>
</tr>
<tr>
<td>City effective at making neighborhoods safer for youth</td>
<td>1.14</td>
<td>2</td>
</tr>
<tr>
<td><strong>School conditions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>City effective at improving schools</td>
<td>1.29**</td>
<td>2</td>
</tr>
<tr>
<td>Local schools good/excellent</td>
<td>0.93</td>
<td>1</td>
</tr>
<tr>
<td>Good/excellent rating of school attended by 5–9 years old in household</td>
<td>0.58**</td>
<td>1</td>
</tr>
<tr>
<td>Good/excellent rating of school attended by youth 10–18 in household</td>
<td>0.96</td>
<td>1</td>
</tr>
<tr>
<td>Youth physically threatened at school in past year</td>
<td>0.81</td>
<td>1</td>
</tr>
<tr>
<td>Youth agrees/strongly agrees she or he feels safe at school</td>
<td>0.98</td>
<td>1</td>
</tr>
</tbody>
</table>

Notes: Data source is the Survey of Adults and Youth (SAY), fielded September-June in 1998–1999 (wave 1), 2001–2002 (wave 2), and 2004–2005 (wave 3) in UHI program and comparison cities. Average approximate sample size per survey wave is 7,400 observations for the UHI city sample and 1,300 observations for the comparison city sample. Sample sizes differ slightly between questions due to missing data and skip patterns. Values represent logistic regression-estimated odds ratio coefficients on terms that interact UHI city residence (vs. comparison city residence) and interviewed in survey Wave 3 (vs. interviewed in an earlier survey wave). Other independent variables in the model included UHI city of residence, interviewed in survey Wave 3, city fixed effects, adult respondent education (less than HS diploma, HS diploma only, and greater than HS diploma), household income (U.S.$50k/year or greater vs. less), adult respondent marital status (married or cohabiting with partner vs. not), adult respondent employment status (employed vs. not), percentage of the year 2000 residential zip code’s households or residents who were under 200% of poverty, who had a college education, who were non-Hispanic black, who were 5 years old or younger, and who were homeowners; and the percentage of dwellings in the zip code that were vacant. Statistical significance levels are labeled as *p < .1, **p < .05, ***p < .01.
also became relatively less concerned about teen pregnancy and vandalism, outcomes that the UHI sites attempted to impact through their strategies to improve youths’ experiences out of school (OR = .82, p < .1 for teen pregnancy and OR = .74, p < .001 for vandalism). Thus, parents’ perceptions were reasonably consistent with the site activities undertaken as a result of the UHI program.

In regard to the long-term outcomes, UHI was successful in producing modest improvements in areas of greatest program activity and in areas in which individual sites focused special attention. Relative to families in the comparison cities, families in the UHI cities came to enjoy greater access to and satisfaction with after-school and other out-of-school services (ORs between 1.30 and 1.46 for availability, and utilization of organized after-school activities and for city effectiveness at developing after-school resources). They also experienced positive changes in their neighborhoods relative to the other distressed cities (OR = 1.24, p < .1, for parent survey respondents viewing their neighborhoods as excellent or good places to raise children).

UHI had positive effects in individual cities where sites made special efforts in a particular area (results not shown). In Richmond, the site made sustained efforts to improve services for young children; day care centers were reported to be good or excellent at significantly higher rates than would be expected, given the comparison city trend (OR = 1.36, p < .05). The Oakland site implemented a multipronged strategy to address violence among kids; subsequently, more youth reported feeling safe in school (OR = 1.47, p < .1), and fewer adults reported unsupervised youth being a big problem in their neighborhood (OR = .75, p < .1). The Philadelphia site pursued major changes in policing strategies, and fewer adults rated “crime and violence” (OR = .69, p < .05) and “not enough police protection” (OR = .63, p < .05) as big problems in their neighborhood. These findings lend credibility to the conclusion that the impacts identified in these analyses were, indeed, the result of program activities.

We tracked 11 indicators using administrative data (see Table 5). The UHI cities showed more positive trends on seven indicators and negative results on four. Outcomes that favored the UHI cities are, as with the survey data, concentrated in the areas of greatest program activity. The UHI cities improved more than comparison cities in decreasing the proportion of births to females younger than 20 years who were already mothers, which may be due to increased efforts, especially in Richmond and Baltimore, to provide nutritional, counseling, and other support services to young first-time and at-risk mothers. The intensive efforts to improve youth safety across the UHI cities may have contributed to the reductions in youth death and arrest.

While survey and administrative data suggest that youth in the UHI cities benefited from the initiative when compared to young people in the comparison cities, caution is warranted. All impacts are small in size and in scope. For example, the average statistically significant program impact is a 5.6 percentage point difference from the comparison city trend, which is approximately a 10% change relative to baseline. And of course in many outcome categories, no statistically significant effects were found. For many indicators, a positive impact is merely a lesser decline; that is, the findings suggest that UHI helped stem further decline rather than resulted in absolute advances. For example, UHI had positive program impacts on the reported effectiveness of cities in developing after-school programs and on the ratings of parks and recreation centers, even though there were small declines in these measures over time, due to the larger declines in these measures in the comparison cities. In addition, while UHI strategies were often targeted at the most disadvantaged communities, statistically significant impacts are found only among better educated parents, suggesting they may have been in a better position to benefit from newly available resources and services.

Judged by the ambitious expectations of its designers, the impacts of UHI fall short. Despite a great deal of hard work and focused effort on activities highly consistent with the program theory, the impacts are in the range of a few percentage points, and in many categories where improvements were anticipated, no such change resulted. Although progress was made in some areas, it was slow and incremental. Dramatic gaps remained between the quality of life for children in these cities and for children in more affluent localities.
The lack of comprehensive change did not appear to be the result of poor implementation on the part of those working in the sites. In fact, in each site changed systems and policies resulted from their efforts, even if particular steps were abandoned or some steps were incompletely realized. Rather, our implementation findings suggest that political, economic, and institutional constraints made it difficult for UHI to realize a dramatic pay off. The narrow, modest results of UHI may reflect a flaw in the logic of the initiative’s design. The question of how much change can result from an intervention of this size and scope, given larger economic and social trends, is critical to the final assessment of UHI. UHI was unusual in its long-term commitment and relatively large expenditures. Yet, the trends it was trying to reverse have persisted for at least 50 years, and the public budgets it was trying to leverage were thousands of times larger than the investment itself. Had it met original

Table 5. UHI and Comparison City Trends and Trend Differences for Selected Administrative Data Measures

<table>
<thead>
<tr>
<th>Data measure</th>
<th>UHI city average</th>
<th>Comparison city average</th>
<th>UHI change minus comparison change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of births with late or no prenatal care</td>
<td>6.0</td>
<td>6.4</td>
<td>−1.2</td>
</tr>
<tr>
<td>Percentage of births that were low birth weight</td>
<td>12.3</td>
<td>10.6</td>
<td>0.4</td>
</tr>
<tr>
<td>Percentage of second or later births, among all births to females younger than 20 years old</td>
<td>25.2</td>
<td>25.3</td>
<td>−0.7</td>
</tr>
<tr>
<td>Percentage of births that were to mothers younger than 20 years old</td>
<td>25.2</td>
<td>25.3</td>
<td>−1.1</td>
</tr>
<tr>
<td>Postneonatal mortality per 1,000 live births</td>
<td>3.1</td>
<td>2.6</td>
<td>0.2</td>
</tr>
<tr>
<td>Mortality due to accidents, violence, or suicide per 100,000 5–9 year olds</td>
<td>20.8</td>
<td>20.0</td>
<td>−0.8</td>
</tr>
<tr>
<td>Mortality due to accidents, violence, or suicide per 100,000 10–18 year olds</td>
<td>34.9</td>
<td>27.1</td>
<td>1.3</td>
</tr>
<tr>
<td>Arrests per 1000 10–17 year olds</td>
<td>8.1</td>
<td>8.2</td>
<td>−0.3</td>
</tr>
<tr>
<td>Gonorrhea cases per 100,000 10–19 year olds</td>
<td>1491</td>
<td>1053</td>
<td>−472</td>
</tr>
<tr>
<td>Attendance rate, grades 1–12</td>
<td>89.9</td>
<td>90.5</td>
<td>0.7</td>
</tr>
<tr>
<td>High school graduation rate</td>
<td>57.9</td>
<td>55.1</td>
<td>1.8</td>
</tr>
</tbody>
</table>

Note: Values are unweighted averages of the five UHI or nine comparison city values. “Late” prenatal care is care that began in the second or third trimester. Postneonatal mortality is mortality from 2–11 months of age. Birth and mortality outcomes were calculated from the National Center for Health Statistics National Vital Statistics System birth and death files. Gonorrhea data are from the Centers for Disease Control STD surveillance system. Attendance data were collected from city school districts and state departments of education. Attendance rates are calculated as aggregate number of days in attendance divided by aggregate number of days of membership. Graduation rates were calculated from the National Center for Education Statistics Common Core data using Christopher Swanson’s formula for the Cumulative Promotion Index (Swanson, 2003). When necessary, denominators were calculated using interpolated census 1990 and 2000 values for 1998 denominators, and American Community Survey 2003 or 2004 population estimates applied to Census 2000 city/age subgroup data for 2003 or 2004 denominators. Due to missing data, the percentage of births with late or no prenatal care in the comparison cities was calculated without Pittsburgh, and the cumulative promotion rate in the comparison cities was calculated without Cleveland.

The lack of comprehensive change did not appear to be the result of poor implementation on the part of those working in the sites. In fact, in each site changed systems and policies resulted from their efforts, even if particular steps were abandoned or some steps were incompletely realized. Rather, our implementation findings suggest that political, economic, and institutional constraints made it difficult for UHI to realize a dramatic pay off. The narrow, modest results of UHI may reflect a flaw in the logic of the initiative’s design. The question of how much change can result from an intervention of this size and scope, given larger economic and social trends, is critical to the final assessment of UHI. UHI was unusual in its long-term commitment and relatively large expenditures. Yet, the trends it was trying to reverse have persisted for at least 50 years, and the public budgets it was trying to leverage were thousands of times larger than the investment itself. Had it met original
expectations, UHI would have been an enormously rewarding investment, the philanthropic world’s equivalent to being a founding investor in Microsoft or Google. However, one can argue that modest impacts in selected and targeted areas are reasonable returns for a cautious investor, if not a venture capitalist.

Politics sometimes stood in the way of achieving UHI’s goals, and weaknesses within providing organizations made the implementation of best practice models, once identified, hard to achieve. Even successful pilot programs were not readily taken to scale because of the complexity of the policy-making process and dramatic shifts in the economic environment. In the earliest stages of UHI, economic prosperity had encouraged optimistic thinking about what could be accomplished; by 2002, the economic reality required the sites to shift their focus in many cases from expanding services to maintaining existing ones. Furthermore, we find no evidence that public funds moved upstream to preventive approaches. And efforts to build consensus and political will had to be frequently repeated as a result of the dynamic political context and turnover among those in elected office.

These obstacles were especially daunting in the arena of education. Efforts to work with the public school systems proved challenging. These systems were under stress from greater scrutiny at the state and local levels and were facing stark financial constraints. As a result, and despite their local plans, the sites abandoned their efforts in this arena. This failure was reflected in our survey data findings. Parents’ assessment of school performance is the only survey indicator that showed statistically significant negative results. Parents in the UHI cities also became far more concerned about school outcomes than would have been predicted by the comparison city trend, which may be related to the UHI sites’ very public and failed attempts to work with the school systems. Administrative data supported the impressions of parents; educational outcomes did not improve, even as other areas, more effectively targeted by UHI activities, did.

Discussion: Strengths and Limitations of Our Integrated Approach

In assessing a program, evaluators ask, “What would have occurred in the absence of that intervention?” In assessing an evaluation design, one might ask a similar question: What would we have concluded about the initiative had we used either a traditional quasi-experimental design or a theory of change approach, rather than integrating the two?

A theory-based approach without a quasi-experimental component would have likely led to the wrong conclusions; UHI would have been deemed a failure. Although the implementation of UHI confirmed much of the underlying theory, the program’s implementation did not come close to following the full causal chain predicted by the theory and the anticipated improvements were rarely achieved. Recall, some program impacts are lesser declines than those in comparison cities rather than absolute improvements. In the end, even an effort as ambitious as UHI, with its large price tag and extended time period, is tiny relative to the larger context in which it operates. Without a comparison group, we would likely have concluded that UHI’s goals were not met, and that this was because the implementation did not fully match the theory. It is only in comparison to similar cities that the impact of UHI can be appropriately identified. In this case, small victories in the UHI sites become more apparent because of the continued declines in other distressed cities. In a theory of change evaluation, it is easy to miss the effects of historical trends that most certainly influence the direction and potential success of CCIs.

Had we relied solely on difference-in-differences analysis, unguided by theory or implementation, our findings would have been more difficult to interpret and would also have been more negative. Had we used a more traditional design, our outcomes would have been chosen early on. Once operationalized and measured, they would not have been revisited or augmented until the
program’s conclusion. Some outcomes, most especially those regarding after-school, might have remained unmeasured throughout the evaluation, despite the fact that this area became salient in the later stages of planning. Furthermore, given the large number of outcomes that were claimed as potential impacts during the design of the program, we would have faced a daunting multiple comparison problem. In the absence of a theory-based evaluation approach, we would have applied strict adjustments for statistical significance, thus guaranteeing that few if any outcomes could be considered impacts. Instead, we chose to be guided by theory and implementation in deciding which outcomes to keep and which to discard.

Thus, had we used either evaluation approach in isolation, the impacts of UHI would have been missed. In the case of UHI, both the theory of change and the quasi-experimental approaches would, by themselves, miss the gains achieved through this effort. The integrated approach implicitly encouraged the recognition of inevitable changes to theory, program, and context in our evaluation. Messy interventions operate in messy worlds; evaluations of CCIs demand a system for accommodating this level of disorder. Our evaluation was expensive, but evaluating complex programs “on the cheap” does not seem to be the answer.

In considering whether, and when, it is worth replicating our integrated evaluation design, it is important to recognize the limitations of this study, which were exacerbated by the complexity and extended time frame of UHI. Impact findings rely on the plausibility of the comparison group as a counterfactual. The comparison cities are closely matched demographically to the UHI cities, but there is room for error. Perhaps the greatest concern is that the UHI sites were, from the start, advantaged over the comparison cities, that our identified impacts reflect the UHI cities being “poised for change.” Of course, given the decades of distress characteristic of these five cities, such change would represent an abrupt and unlikely turn of events. Furthermore, neither our qualitative nor quantitative findings support the idea that the UHI cities were advantaged at their point of selection. Rather than dismissing the possibility of using quasi-experimental designs for the evaluation of place-based efforts, the field would benefit from further discussion of tools for selecting appropriate comparisons. We believe our work has moved this discussion forward, but others will find its flaws and make their own improvements.

The findings are also limited by the nature of the data we were able to collect. Even with the use of multiple methods of data collection and a flexible approach, there were critical service areas in which no single source of administrative or fiscal data exists, and even multiple data sources were difficult to disaggregate and reaggregate to provide a comprehensive picture of services and expenditures. UHI site staff members were repeatedly challenged to monitor activities and outcomes in areas in which the scarcity of data required primary data collection or “best guesses” from existing data sources (Silver et al., 2002; Weitzman, Silver, & Brazill, 2006). Furthermore, administrative data are available only with a considerable time lag; to the degree that the greatest results may be evident in later years, they are not incorporated into the evaluation. The lack of existing administrative data and the tremendous costs associated with primary data collection are significant obstacles to an integrated design, where evaluation resources must be split between time spent in the field doing qualitative, theory-based implementation research and time spent collecting and analyzing survey, administrative, and fiscal data.

As mentioned earlier, our integrated design stressed evaluating the national initiative rather than its local components. RWJF provided funds to allow the sites to have their site-specific research and evaluation needs addressed locally. This evaluation activity was never intended to serve as the local arm of the national evaluation, as is often done in multisite interventions. Thus, our national evaluation is limited in its ability to identify the specific activities within any given site that were most beneficial to that site’s aims.

Finally, the UHI evaluation augmented existing data by developing and fielding the SAY survey. Nevertheless, survey data have their limitations. For example, many of the findings reported here are derived from self-reports of perceptions of problems, conditions, and institutional performance, not
direct measures of those problems, conditions, or institutions. This, however, may not be a signif-
icant problem for the evaluation of CCIs because they rely heavily on community engagement,
changes in norms and attitudes, and local understanding of local conditions; perceptions are at the
core. Of course, survey- and sample-based measures are subject to sampling and measurement error,
as well as response bias. We attempted to control for these problems in the best ways possible.

We do not mean to suggest that our evaluation has been without flaws nor would we argue that
our integrated approach is appropriate to, or desirable for, the evaluation of all CCIs or ABIs. This
was an expensive and labor intensive study requiring a team of evaluators with skills cutting across
disciplines. As always, the level of investment in evaluation should be commensurate with the
investment in the program. That said, it is important to recognize that comprehensive place-based
interventions have become popular despite the lack of solid evidence regarding their impacts. Our
study indicates that it is possible to begin to estimate the impacts of these interventions; combining
these estimates with strong information about context and the process of change will allow decision
makers to rethink their expectations for CCIs and to refine their models to achieve greater impact.

Conclusion

Philanthropies like RWJF take bold steps in trying new approaches to solving old problems. In trying
to evaluate these new approaches, evaluators are taken to the limits of their field’s capacity to dis-
tinguish program effects from the myriad changes in a community over time. Given this formidable
task, evaluators should use every tool at their disposal and use them in new and synergistic ways. In
developing and using an integrated approach, we attempted to meet the challenge of evaluating a
messy initiative by bridging the gap between qualitative and quantitative methods and between the-
ory of change and quasi-experimental designs. As this article demonstrates, evaluations of complex
initiatives like UHI can be strengthened by narrowing this divide.

Authors’ Note:
The authors thank members of the Advisory Committee to the National Evaluation of the Urban
Health Initiative (UHI) for thoughtful advice. In addition, the authors thank James Knickman for his
ongoing support of these activities and Martha Holleman for critical insight and assistance. The
cooperation and insights offered by the staff and board members at each of the UHI sites was essen-
tial to this work. An earlier version of this article was presented at the Annual Meeting of APPAM,
November 2006, where we received helpful suggestions from Dahlia Remler. The ideas, opinions
and conclusions expressed within are the authors’ alone. This study received approval from the Uni-
versity Committee on Activities Involving Human Subjects at New York University.

Declaration of Conflicting Interests
The authors declared no conflicts of interest with respect to the authorship and/or publication of this
article.

Funding
Funding for this evaluation was provided by the Robert Wood Johnson Foundation.

Note
1. OR = odds ratio, representing the observed odds of the outcome divided by the predicted odds of the out-
come estimated from the difference in odds between program and comparison cities in the baseline period,
and the difference between baseline and post-intervention odds in the comparison cities, adjusted for covariates. Reported impacts are significant at the .1 level or lower.

References


