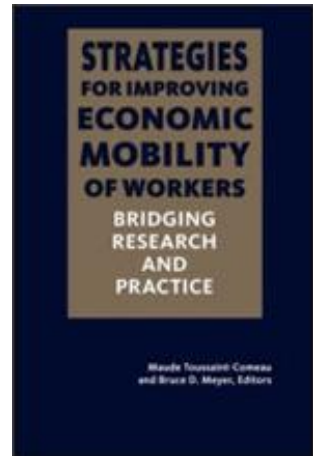

Upjohn Institute Press

Comparing Apples to Oranges When Evaluating Community- Based Programs and Services

Robert John LaLonde
University of Chicago



Chapter 12 (pp. 197-208) in:

**Strategies for Improving Economic Mobility of Workers: Bridging
Research and Practice**

Maude Toussaint-Comeau, Bruce D. Meyer, eds.

Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 2009

DOI: 10.17848/9781441631992.ch12

12

Comparing Apples to Oranges when Evaluating Community-Based Programs and Services

Robert J. LaLonde
University of Chicago

Community organizations have proliferated over time. These non-profit entities may provide a range of services, or they may provide a very specific set of services targeted to residents of a geographically defined area, such as a neighborhood, a school community, a parish, or even a city or town. Their services include, but are not limited to,

- Prisoner reentry
- Community policing
- Job placement
- School tutoring
- Job training for new immigrants
- Home visitation for new mothers
- Drug treatment and drug treatment referrals
- Domestic violence counseling and centers
- Early childhood education initiatives
- Programs to encourage minority arts participation
- Small business assistance for low-income persons
- Food banks and services for the homeless.

These services have been provided by both nonsectarian and faith-based organizations. They are designed to improve outcomes for recipients of the services, for their families, and for the communities in which they live.

Financing for these services usually comes from charitable contributions, from members of the community organization providing the service, and from outsiders, but it can also come from grants from foun-

dations. Each year these foundations provide community organizations with billions of dollars of support through grants.¹

In recent years foundations have increasingly asked community organizations to evaluate the programs and services that have been at least partially financed by their giving. Although it is not always clear what these foundations have in mind when they ask their grantees for these evaluations, and sometimes what they ask for is inappropriate or infeasible, they are correct to ask their grantees to be quantitatively more accountable than they have been in the past for the dollars that they have spent.

In this chapter, I explain why evaluations often do not constitute a cost-effective use of foundation resources, nor do they constitute a productive use of the time and resources for the community organizations that receive the grants. This point holds especially true for “impact” or “cost-effectiveness” evaluations. In addition to the list of services given above, empowerment or enterprise zones, tax increment financing methods (TIFs), recycling programs, and “NIMBY” (“not in my backyard”) disputes over such issues as the siting of a transfer station or a power generating plant also can fall under this rubric of community-level services. One of my conclusions here is that despite this varied list of services and policies, the challenges associated with evaluating them are virtually identical.

These challenges include not enough baseline information to complete a timely and rigorous evaluation, too little quantitative information on the services provided, who receives them, and how the services delivered differ from those intended. At other times the size of the intervention is relatively small and its impacts could only be detected with an extremely large sample of participants.

However, even when impact evaluations or cost-effectiveness evaluations are inappropriate, there is usually other valuable information about the community organizations’ practices and performances that the organizations can collect and quantify. This information is valuable, not only to the foundations and other nonprofit and public-sector organizations that fund these community organizations’ services, but also to their stakeholders and the organizations themselves as a tool to improve their program operations.

WHAT TO EVALUATE?

Should evaluators evaluate for community organizations 1) the impact of grants that they receive on the services that they provide to their clients or 2) the impact of their services on participants' outcomes? In the former case, suppose a foundation asks its grantee to assess the impact of the foundation's funding on the quantity or quality of services it provides to members of its community. In the second case, suppose a foundation asks its grantee to evaluate the impact of the community organization's primary services on the residents it serves and on who receives services.

The Impact of Grants on a Single Organization's Performance Cannot Be Evaluated

The first of these objectives—to evaluate the effects of the foundation's funding on the services that a single community organization provides—is not feasible. Although foundations would like to know what difference their giving has made to the organizations that have received the grants, and the organizations that received them would like to tell all the foundations that support their programs how important their funding was to them, yet without including often implausible assumptions into the analysis it is not possible to evaluate this question for a single community organization. In practice, there are many factors affecting the quality and quantity of services provided by a community organization. Besides the foundation's grant, the community organization's services also are affected by grants from other foundations and from the public sector, hiring and departures of personnel, and changing social and economic factors in the community.

To see the difficulty associated with the task of evaluating how a foundation's funding affects a community organization's services and performance, I manipulate the terms in a simple framework. First I define some notation:

A = a measure of services provided by the community organization *after* it received funding from the foundation.

A' = a hypothetical measure of services that would have been provided by the community organization if its proposal for funding from the foundation had been rejected.

A_{-1} = a measure of services provided by the community organization *before* it received funding from the foundation.

B = a measure of services provided by another community organization whose proposal for funding from the foundation had been rejected.

D = a measure of services provided by another community organization that never even applied for funding from the foundation.

In order to assess whether its grants have made a difference, a foundation may ask community organizations to provide it with a measure of the following impact:

$$\text{IMPACT} = A - A'$$

This impact measures the difference between the quantity (or quality) of services provided after the community organization received the foundation's grant (A) and the quantity (or quality) of services that would have been provided if the community organization had not received the foundation's grant (A'). Notice that measure A' is not the same as the measure of services provided by the community organization before it received funding from the foundation (i.e., A_{-1}).

Put this way, the problem is readily apparent. Although the term A can be measured, the term A' cannot. To measure A' requires literally turning back the clock to the point at which the foundation was making its funding decisions (Holland 1986). Once events have been theoretically returned to that time, the foundation then rejects the community organization's proposal and term A' measures the services provided under this alternative scenario.

Program evaluators refer to this problem of not being able to turn back the clock as "the evaluation problem" (Heckman, LaLonde and Smith 1999). Since turning back time is impossible, evaluators must fill in these missing data with an estimate of measure A' . In one approach,

to estimate A' , evaluators use a measure from a similar community organization that did not receive funding from the foundation. One possibility in this approach is to use information from another community organization whose proposal for funding was rejected by the foundation (denoted by B), as follows:

$$(1) \quad \text{IMPACT} = A - B .$$

Another possibility is to use information from another community organization that never applied for funding (denoted by D), as follows:

$$(2) \quad \text{IMPACT} = A - D .$$

Finally, a third possibility is to use information from the community organization before it received funding from the foundation (denoted by A_{-1}), as follows:

$$(3) \quad \text{IMPACT} = A - A_{-1} .$$

Immediately, it is apparent that we have at least three different estimators of the impact of the foundation's grant-making, and none of them are necessarily equal to the true impact of the foundation's grant-making, which is $A - A'$.

All three of the foregoing impact measures have the same shortcoming: namely, that measures B , D , and A_{-1} are not the same as measure A' . They are oranges, and measure A' is the apple. Consider the following rearrangement of terms in the three impact measures:²

$$(1') \quad \text{IMPACT} = A - B = (A - A') + (A' - B) ;$$

$$(2') \quad \text{IMPACT} = A - D = (A - A') + (A' - D) ;$$

$$(3') \quad \text{IMPACT} = A - A_{-1} = (A - A') + (A' - A_{-1}) .$$

The implication of these three estimators of the impact of a foundation's giving on a (single) community organization's services or performance is that evaluators can reliably and precisely estimate the counterfactual value A' using measures based on data from other community

organizations or from the same community organization at some point before it was funded by the foundation. Experience indicates that neither of these possibilities works well in practice.

To understand why not, consider the first of these measures: $A - B$. Information drawn for measure B is from another community organization that applied for funding but whose proposal was rejected by the foundation. As a result, this rejected community organization is one that the foundation has looked at carefully and decided that its proposal and its likely performance were inferior to those of the community organization it decided to fund. It would be remarkable if the counterfactual measure A' for the funded community organization would be well approximated by measure B from another community organization whose application for funding was rejected.

Foundations' funding decisions are carefully made. Foundations fund proposals that best serve their objectives and mission. So it is unlikely that the set of funded community organizations would be like their unfunded counterparts. However, suppose for the sake of argument that foundations' funding decisions were not carefully made. Even in the extreme and unlikely case where the foundation did not deliberate over the proposals it received and instead flipped a coin to determine funding, it would be likely that the impact measure of $A - B$ is in error. To be sure, the coin flip ensures that *on average* the impact measured by $A - B$ approximates the true impact measured by $A - A'$. However, this point only holds on average, not for any single comparison between a funded community organization and one of its rejected counterparts.

The problem with using even a randomly rejected community organization's measure of services or performance, B , so as to use $A - B$ to estimate the true impact of the foundation's funding, $A - A'$, is matching error. Even in the absence of the foundation's funding, these community organizations are not the same, and as a consequence their measures of performance also should not be the same. The difference $A' - B$ contains both 1) the performance differences between the community organizations and 2) the errors associated with comparing or matching organizations that are not the same. These expressions tell us that in order to measure the impact of a foundation's giving on a community organization's services and performance, it is necessary that we compare measures for organizations that are the same or essentially the same.

This point raises another question: since other community organizations are not the same, could evaluators use the information on the same community organization before it was funded by the foundation? The problem here, as I explained above, is that many factors affect an organization's performance—including different personnel, funding from other nonprofit organizations, changing economic and social factors within the targeted communities, even the entry or emergence of another community organization in the area. Even here there is “matching error.” An organization in the past is not the same organization as it is constituted in the present.

Evaluate the Average Impact of Funding on a Group of Organizations' Performances

The discussion in the preceding section explains why evaluators contend that we cannot evaluate the impact of an intervention—such as a foundation's giving—on the performance of a single organization. This principle also extends to individuals, and for the same reasons. There are many influences on an individual's outcomes besides those associated with the services that they receive from a community organization. A comparison between an individual who receives services from a community organization and another individual who does not receive such services usually tells us more about preexisting differences between these two people than it does about the differences in effectiveness of the services themselves.

The standout solution to what I described above as the matching-error problem is to evaluate the average impact of a foundation's giving on many different, but similar, community organizations by comparing their measures of performance to those of many other community organizations that have not received funding from the foundation. In this case, the expectation is that if the evaluators have done a good job of matching these other, unfunded community organizations to the funded community organizations, then the matching errors will average out.

The implication of this proposed solution to the evaluation problem is that it is not feasible to evaluate the performance of a set of, say, four or five community organizations that have been supported by a foundation. However, it may be feasible to evaluate the impact of a foundation's grant-giving by evaluating the average impact on a set of, say,

100 similar community organizations. In this latter case, it is possible that the matching errors average out, leaving a credible and possibly precise estimate of the average impact across the 100 organizations that the foundation chooses to fund. The underlying assumption is that the matching errors average out in large samples of community organizations or of service recipients.

In more colloquial terms, under some circumstances it may be an effective evaluation strategy to compare large numbers of apples to large numbers of oranges. To be sure, when evaluating the impact of foundation giving on a few community organizations' performances, we are faced with the familiar intractable problem of comparing apples to oranges. But when evaluating the impact of foundation giving on a large number of organizations, it is possible that the matching errors associated with comparing apples to oranges will average out. One setting in which they are likely to average out occurs under the hypothetical but unrealistic scenario that foundations' funding decisions are made randomly from their pools of applicants.

In practice, this proposed statistical solution does not help most foundations and nonprofits evaluate their giving, because of their practice of targeting their resources to a few community organizations with well-conceived proposals for providing innovative services. These foundations' giving policies are likely good ones. Yet despite this, in the vast majority of cases, it makes little sense for the foundations to insist that community organizations evaluate the impact of their funding.

Evaluate Services Regardless of the Source of Funding

The preceding discussion does not imply that foundations should forgo evaluation of important services targeted toward communities and individuals. Foundations can improve their own funding decisions and provide valuable information that will help community organizations operate their programs and deliver their services if they evaluate the services offered or the services received and not the organizations themselves.

Therefore, instead of asking, "What difference did our funding make to a particular organization and by extension to the organization's 'clients'?" foundations should ask instead, "What impact do these particular services have on individuals who receive them?" Such evaluations

should focus on services and the individuals who received them and not on the improved performance of organizations that resulted from the funding provided by the foundation.

Even when taking this approach, it is not always an efficacious use of resources to insist on an impact evaluation. Prior to the impact evaluation, there are at least three other steps that foundation program officers and community organizations should take:

- 1) Collect baseline and programmatic information on participants,
- 2) Implement a study to see if the services are delivered as intended, and
- 3) Assess the size of the intervention and its theoretical impacts.

In most cases community organizations' data collection efforts have lagged behind other important work that the organizations do. Since evaluation is a quantitative endeavor, outcomes of interest as well as services provided must be measured before any formal evaluation occurs.³ The key point is that if the outcome cannot be measured, it cannot be evaluated.⁴

There is no point in evaluating a program or service before we learn how it operates in practice. An implementation study must precede an impact evaluation. Otherwise, even if the impact evaluation demonstrated that the service was effective, the potential for the service to be replicated would remain uncertain, because it would not be clear how the program actually operated in practice in the field. An implementation study should, among other things, document whether the services offered and received differ sufficiently from what is intended. Such differences arise because of the organization's or service provider's performance, because of a mismatch between the services and the recipients, or because resources are insufficient to implement the intended design. Before initiating an impact evaluation, such implementation questions need to be resolved.

Finally, the size of an intervention may be too small for its impact to be detected with a modest sample of recipients. For example, a 20-minute counseling session on how to find a job could be cost-effective, because the intervention is so inexpensive on a per-person basis. In this situation, the impacts would not need to be large for these services to be cost-effective.⁵ However, an outcome such as employment rates or unemployment durations could be sufficiently variable among mem-

bers of the community that a large sample of service recipients would be required in order to detect small impacts but cost-effective services.

These challenges associated with evaluating community organizations and their services are not unique and arise in other similar circumstances. For example, in evaluating enterprise zones, studies have addressed the evaluation problem by comparing economic activity in areas after they have been designated as enterprise zones to the economic activity prior to their receiving this designation.

A second approach compares these zones to areas with similar characteristics (Greenbaum and Engberg 2000; Holland 1986). Using both approaches, these areas have been defined as communities, census tracts, neighborhoods, or cities. In this alternative approach, evaluations attempt to construct a quasi-experimental setting in which areas that are designated as enterprise zones are compared to observationally similar areas that are not designated as enterprise zones.

No matter what approach is used, these evaluations cannot produce reliable estimates of the impact of a single or even a few enterprise zones on community and individual outcomes. In order for such evaluation studies to reliably measure impacts of enterprise zones, they must carefully measure outcomes and area characteristics for a sufficiently large number of these zones and their corresponding comparison areas. In this way, they measure the average impact of the enterprise zone strategy for a sample or particular "population" of communities.

CONCLUSION

High-quality evaluations are costly, and they are cost-effective only if they lead to some significant outcome. Saving an effective program or set of services that would otherwise be eliminated is a worthy purpose of evaluation. So is documenting socially significant net impacts for a new, innovative program.

However, in many instances involving community organizations, the goals of evaluation should be more modest. Often they should not focus on impact evaluation or cost-effectiveness, but on simply measuring and collecting data on program services. Providing operators and foundations with this information alone can provide valuable insight

into how the program operates or how services are delivered and the challenges that must be overcome in order to affect recipients' outcomes. At the very least this information can improve program management. As I have explained above, this information also is essential for considering whether it is a good use of resources to initiate an impact evaluation of these programs and services.

Notes

1. See <http://foundationcenter.org/findfunders/statistics/grantsampling.html> for statistics on foundations believed to account for about half of all grant-making in the United States. Total grant giving from foundations for all purposes is thought to have amounted to about \$40 billion in 2006.
2. In each case in Equations (1') through (3'), I simply subtract A' and add A' to Equations (1) through (3), above.
3. An example in Chicago is the study titled *Mapping Cultural Participation in Chicago* (LaLonde et al. 2006), funded by the Joyce Foundation. Originally, the foundation asked researchers at the National Opinion Research Center (NORC) and the Irving B. Harris Graduate School of Public Policy Studies to evaluate the effectiveness of various initiatives designed to promote minority participation in the large Chicago-area arts organizations. The researchers countered that without baseline information in hand such an evaluation would not succeed. So instead the foundation and the researchers agreed to collect baseline information on participation in the arts by individuals living in the Chicago metropolitan area. See LaLonde et al. (2006).
4. To address the costs of data collection and data management, smaller community organizations should consider hiring as interns the very talented students from the select high schools in their areas for data collection purposes. Many of these students have quantitative and computer skills that can be very useful to a standard community organization. They could be hired at very low cost.
5. An example of a small-sized intervention—a letter that unemployed job seekers were instructed to show potential employers—is discussed in Burtless (1985).

References

- Burtless, Gary. 1985. "Are Targeted Wage Subsidies Harmful? Evidence from a Wage-Voucher Experiment." *Industrial and Labor Relations Review* 39(1): 105–114.
- Greenbaum, Robert, and John Engberg. 2000. "An Evaluation of State Enterprise Zone Policies." *Policy Studies Review* 17(2/3): 29–46.
- Heckman, James J., Robert J. LaLonde, and Jeffrey A. Smith. 1999. "The Economics and Econometrics of Active Labor Market Programs." In *Handbook of Labor Economics*, Orley C. Ashenfelter and David Card, eds. Vol. 3A. Handbooks in Economics 5. Amsterdam: North-Holland, pp. 1865–2084.
- Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81(396): 945–960.
- LaLonde, Robert J., Colm O'Muircheartaigh, Julia Perkins, Ned English, Diane Grams, and Carroll Joynes. 2006. *Mapping Cultural Participation in Chicago*. Monograph. Chicago: Cultural Policy Center, University of Chicago.

Strategies for Improving Economic Mobility of Workers

Bridging Research and Practice

Maude Toussaint-Comeau
Bruce D. Meyer
Editors

2009

W.E. Upjohn Institute for Employment Research
Kalamazoo, Michigan

Library of Congress Cataloging-in-Publication Data

Strategies for improving economic mobility of workers : bridging research and practice / Maude Toussaint-Comeau and Bruce D. Meyer, editors.

p. cm.

Includes bibliographical references and index.

ISBN-13: 978-0-88099-352-4 (pbk : alk. paper)

ISBN-10: 0-88099-352-9 (pbk : alk. paper)

ISBN-13: 978-0-88099-353-1 (hardcover : alk. paper)

ISBN-10: 0-88099-353-7 (hardcover : alk. paper)

1. Migration, Internal—United States. 2. Occupational retraining—United States. 3. Housing subsidies—United States. I. Toussaint-Comeau, Maude. II. Meyer, Bruce D.

HB1965.S76 2009

331.12'70973—dc22

2009022482

© 2009

W.E. Upjohn Institute for Employment Research
300 S. Westnedge Avenue
Kalamazoo, Michigan 49007-4686

The facts presented in this study and the observations and viewpoints expressed are the sole responsibility of the authors. They do not necessarily represent positions of the W.E. Upjohn Institute for Employment Research.

Cover design by Alcorn Publication Design.

Index prepared by Diane Worden.

Printed in the United States of America.

Printed on recycled paper.