Evaluation of Community-Wide Interventions

Austin Nichols

July 2013

Policymakers often turn to place-based initiatives to address complicated and persistent community-wide problems such as concentrated poverty and lack of access to quality schooling, jobs, and health care. Most place-based initiatives are fundamentally incompatible with commonly used evaluation methods, and random assignment frequently represents a very poor evaluation design in this setting. However, there are evaluation methods that can address these challenges.

Most place-based interventions are (1) designed to capitalize on spillover effects from participants in programs on those who never participate; (2) offer linked sets of services and case management rather than a single, specific treatment or intervention and incorporate continuous improvement models where the treatments are continually altered in response to ongoing data collection on how outcomes respond; and (3) are tailored to specific local conditions and capacity in a way that makes each initiative unique in its particular mix of program components and how those match conditions on the ground in the initiative's location.

Any one of these three properties would preclude the use of random assignment experiments, or related quasi-experimental methods based on thought experiments, but together they make any type of evaluation challenging.

Innovative methods, many of which use propensity score reweighting methods in novel ways, can address the fundamental challenges of spillover, continual improvement in treatments, and limits to generalizability due to place- and population-specific designs.

Spillovers

Smith (2011) points out many of the challenges in evaluating a place-based initiative, such as Choice or Promise Neighborhoods, but argues that several traditional methods of analysis can be used to evaluate program components:

> an experimental design could be used in Choice or Promise Neighborhoods to evaluate the impact of a specific service... if a natural experiment was identified, perhaps through a lottery to receive a specific service (leaving a treatment group of those who gained access to the service and a control group of those who did not).

Such an experimental model also underlies quasi-experimental estimates of the impact of a program component, but these are not the types of impacts a saturation model of service delivery aims to achieve. The goal of most place-based initiatives is that those who do not receive services are affected by services delivered to others, so the obvious control group is never a valid control group.

The framework of potential outcomes (Holland 1986) that justifies experimental trials and motivates quasi-experimental designs (Nichols 2007) does not apply if how widely a treatment is applied has an impact on the effectiveness of treatment, or if some units' treatment status affects others. Think of an experiment to measure the average effect of a vaccine: if 20 percent of a community gets a flu shot, it will be less effective even for that 20 percent than had 80 percent gotten the shot. Similarly, a program to increase reading or math proficiency randomly assigned to half the children in a classroom also affects the control group, if there are peer effects.

Rubin (1986) advocates changing the unit of analysis if the "no spillovers" assumption fails, for example comparing across classrooms or schools (either with all children assigned to the same treatment status, or with different numbers of children assigned per unit) instead of children. But we cannot easily randomly assign neighborhoods to receive a standardized treatment in order to measure the effect of a community-made, culturally specific, place-based initiative.

There is an alternative approach that can address spillover effects, requiring us to construct synthetic control group communities to represent what would have happened in an area if a place-based initiative had not taken root. Abadie and Gardeazabal (2003) use a weighted average of other regions in Spain to represent the counterfactual Basque region without terrorism. Abadie, Diamond, and Hainmueller (2010) found that tobacco use fell relative to a synthetic control region representing California without new restrictions. Synthetic controls or matched comparison sites are viable methods for causal inference in the case of interventions that take place at the level of a community, not an individual person or family. But we must remember that there may be a nonlinear response to dose when spillovers are present, and measure dose accordingly.

Linked services and continual improvement

A common feature of place-based initiatives is the linking of services via a case management system designed to ensure that participants do not fall through the cracks. The notion is that coordinated services are more effective, that the sum effect of coordinated services is greater than the summed effects of individual uncoordinated services. That is, even if the control group were not affected by treatment saturation in a community, it would be inappropriate to run a random assignment evaluation on each service, because a bundle of disconnected services is not the treatment to be examined.

Even if we accept that Abadie's synthetic control group approach can solve many of the evaluation problems arising from the spillover effect from treated to untreated families in a community, the design of many place-based initiative involves watching outcomes closely and altering the treatment to continually improve the effectiveness of services. The outcomes at each stage can thus affect the type of treatment delivered. There is no one treatment or regimen of treatments to be evaluated, but rather a philosophy of results accountability and a large menu of treatments to be applied as needed.

A set of evaluation methods known as g-estimation (Robins 1997; Witteman et al. 1998) can account for treatments that respond to observed intermediate outcomes, such as antiretroviral therapy treatments that are adjusted in response to cell counts (Robins et al. 1992) or educational interventions that occur in response to achievement setbacks. Under some circumstances, a propensity score reweighting technique produces good estimates (Robins, Hernán, and Brumback 2000). The parametric g-formula is another approach described in Taubman and colleagues (2009).

Measuring the impact of a therapy that turns on in response to a poor outcome, measured at certain points in time using a variety of traditional quasi-experimental methods, can give the wrong answer, because the treatment status is affected by current indicators of sickness, which in turn are affected by past treatment. At each point in time, those treated are worse off than those not treated, and it's hard to control away the difference in their underlying vulnerability. This is common in social policy settings, for example, in education or income maintenance, where a student who is behind is treated with increased teaching resources, or a food-insecure family is treated with food assistance.

One of the central problems in applying this kind of method is that we need very good data on both the nature of the treatment being applied at each point in time and the indicators that drive treatment decisions. In the education example, we need good data on the in-school and out-of-school instruction given to a failing student, as well as the scores on assignments that motivate different levels of intervention. Only a few place-based interventions are capable of supplying this kind of data, but Promise Neighborhoods may come close to this high standard, if partners collect detailed data on ongoing treatments and indicators of outcomes as planned.

The importance of place and population

The setting of any treatment matters. This is true even where the treatment is generic or of minimal value, such as a placebo sugar pill made by a drug company. (Numerous studies have shown health benefits of placebos administered in a medical setting, and new research efforts are exploring the factors that affect the measured effectiveness of placebos.) This is even more true when the intervention is a complex set of programs with complicated linkages across programs. The same program will have different effects in different places, depending on what other programs are in operation, as well as in different populations. But a program with the same design and the same name will not be the same program in a different place; implementation analysis is required to understand the nature of the treatment and how it interacts with the characteristics of its setting.

Many characteristics of places can affect how a program delivers services and what its effects are on different populations. Classic work by Kain (1968) and Wilson (1987) motivates much of the focus on place-based interventions by pointing out the impacts of concentrated disadvantage. These impacts may not affect all groups equally, and hypotheses about spatial mismatch and tipping points are all subject to debate, but the influence of the topology of a place in terms of geographic features, transportation, and job and social networks, is important to bear in mind in any discussion comparing across places.

When measuring the effectiveness of place-based initiatives, it's challenging to know how to draw inferences from comparisons that best pool findings across sites. Traditional meta-analytic regression techniques require both that the interventions be essentially identical except for a few characteristics measured without error by simple variables and that individual estimates be unbiased. When different sites provide different kinds of evidence, a simple meta-analytic regression can produce badly biased inference about average effectiveness.

Many community change initiatives have both transformed the community and displaced its residents. For example, HOPE VI, a federal program to transform blighted public housing, was very successful in improving housing, reducing crime and poverty, and seems to have produced positive spillovers (Popkin et al. 2004), but it also displaced many of the original residents. If one improves the statistics on children in a neighborhood by evicting a disadvantaged population and moving in a well-off group of children, we have no evidence of improvements in outcomes. More subtly, if only residents who are the most resilient and capable of navigating a changing community environment stay during a period of neighborhood revitalization, we will obtain a biased estimate of effects by comparing their outcomes before and after an intervention.

Similarly, if half the target group that is offered an intervention takes it up, measuring effects for only that group gives a biased view of the effects that might have been achieved had the intervention achieved a higher take-up rate. Even if we ignored spillover effects and ran a random assignment evaluation, an instrumental variables approach that estimates the impact of a treatment on the treated, or an intention-to-treat analysis that averages the impact of treatment on the treated with the minimal impacts of treatment on those who were offered treatment and turned it down, both miss measuring the impact that could have been achieved with continually improved service delivery increasing take-up and changing the population receiving treatment.

Evaluators need to be clear and intentional about a choice of the target population of an actual intervention, and to be disciplined about maintaining the focus on that target population during an evaluation and collecting the data required to answer questions about said population, even if they are moving during the intervention. We have to not only consider the treatment as implemented, but also think about the larger population to which we wish to generalize.

Illustrative examples

Perhaps the most salient place-based initiatives are Choice Neighborhoods, Promise Neighborhoods, and the Housing Opportunity and Services Together demonstration. The last is more amenable to traditional evaluation methods than some other place-based initiatives, but "because adding comparison groups is contingent on gaining additional funding, the Urban Institute may not be able to measure program impacts in a traditional sense" (Popkin et al. 2012). Choice will be evaluated once sufficient data have accumulated, with the evaluation method still to be designed, but the federal Promise Neighborhoods initiative currently has no evaluation scheduled.

Past evaluations have pointed out the synergy possible in a locally tailored place-based initiative, but we do not have conclusive causal evidence on the higher effectiveness of coordinated services in these settings, or the extent of spillover effects on nonparticipants. Still, case management is very promising.

> The Chicago Family Case Management Demonstration was ... from March 2007 to March 2010 ... remarkably successful in implementing a wraparound supportive service model for vulnerable public housing residents.... Strikingly, participants reported gains in employment, health, improved housing and neighborhood conditions, and reduced levels of fear and anxiety.... The additional costs for the intensive services were relatively modest, suggesting that it would be feasible to take a carefully targeted intensive service model to scale. (Popkin et al. 2010, 2)

Lessons learned from prior (non-random assignment) evaluations of place-based initiatives to effect community change are summarized in Kubisch and colleagues (2010). In a tremendously useful case study of Making Connections, Fiester (2011) discusses the unique challenges of place-based research, emphasizing how defining and measuring the treatment can affect services and evaluation.

The most prominent random assignment study of a place-based initiative was Moving to Opportunity, where public housing residents were offered vouchers to move to lower poverty areas, but relatively few took the vouchers. Those who did often did not stay long, and the control group was largely displaced by transformations of public housing undertaken in HOPE VI (Turner, Nichols, and Comey 2012). So not only were control group members subject to some spillovers, as their friends moved to better neighborhoods, but the control group also got a large unintentional dose of another type of treatment. Geographic spillover played a role in the assessment of the effect of HOPE VI on crime (Cahill, Lowry, and Downey 2011). Unplanned treatments affecting controls plays a role in many designs using matched comparison sites, such as evaluations of the Community Healthy Marriage Initiative (Bir et al. 2012) and Safe City (La Vigne, Owens, and Lowry 2010), and can also plague random assignment studies.

Equally challenging to evaluate are demonstrations implemented under the 1996 Moving To Work public housing demonstration, such as Jobs-Plus and time limits on housing assistance. In the case of Jobs-Plus, a random assignment evaluation (Bloom, Riccio, and Verma 2005) chose one housing development per site to implement a package of job search assistance, vocational training, rent modification to improve returns to work, and peer-to-peer community building. Of six sites, only four implemented the whole package, and results were driven by three of those four. However, we do not know to what extent increased work by those in treatment sites affected the work of residents in the control sites, so randomization cannot guarantee unbiased answers. Because there are essentially six observations with three positive results, statistical judgments are also difficult at best. Further, the increased earnings reflected in unemployment insurance earnings records could simply be the result of workers moving from the gray economy into wage and salary jobs.

Takeaways

Place-based initiatives cannot be evaluated by randomly assigning a fraction of potential participants to get a single service and a fraction to not get that service. These initiatives aim to affect the control group of such an experiment nearly as much as the treatment group; spillovers are the goal, not a byproduct (Garfinkel Manski, and Michalopoulos 1990). Also, most place-based initiatives are more than the sum of their parts. Even collecting detailed implementation information about which services are offered and how well they serve their target clientele does not tell us about the interconnections across services that can improve the long-term effectiveness of each service.

A credible evaluation of place-based initiatives uses an evaluation design that allows for spillovers (which a standard random assignment approach does not), accounts for dynamic adjustments of treatment to intermediate outcomes, and correctly models the nature of the treatment being studied. Describing the treatment requires knowing details about implementation on the ground, and comparing the results of treatment in vastly different places is difficult at best. However, a synthetic control approach married to a g-estimation design can address a single site's effect on outcomes, and a meta-analytic approach that uses propensity score methods to adjust for differences in population served holds out the hope of aggregating information across sites.

References

Abadie, Alberto, and J. Gardeazabal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review* 93(1): 113–32.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105(490): 493–505.

Bir, Anupa, Robert I. Lerman, Heather Kofke-Egger, Austin Nichols, and Kevin Smith. 2012. *The Community Healthy Marriage Initiative Evaluation: Impacts of a Community Approach to Strengthening Families Technical Report.* OPRE Report 2012–34B. Washington, DC: US Department of Health and Human Services.

Bloom, Howard, James A. Riccio, and Nandita Verma 2005. *Promoting Work in Public Housing: The Effectiveness of Jobs-Plus.* New York: MDRC.

Cahill, Meagan, Samantha S. Lowry, and P. Mitchell Downey. 2011. "Movin' Out: Crime and HUD's HOPE VI Initiative." Washington, DC: The Urban Institute. http://www.urban.org/publications/412385.html.

Fiester, Leila. 2011. *Measuring Change While Changing Measures: Learning In, and From, the Evaluation of Making Connections.* Baltimore, MD: The Annie E. Casey Foundation.

Garfinkel, Irwin, Charles E Manski, and Charles Michalopoulos. 1990. "Micro Experiments and Macro Effects." Discussion Paper 922–90. Madison, WI: Institute for Research on Poverty. http://www.irp.wisc.edu/publications/dps/pdfs/dp92290.pdf .

Holland, Paul. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81(396): 945–60.

Kain, John. 1968. "Housing Segregation, Negro Employment, and Metropolitan Decentralization." *Quarterly Journal of Economics* 82(2): 175–97.

Kubisch, Anne C., Patricia Auspos, Prudence Brown, and Tom Dewar. 2010. *Voices from the Field III: Lessons and Challenges from Two Decades of CCI Efforts.* Washington, DC: The Aspen Institute.

La Vigne, Nancy G., Colleen Owens, and Samantha S. Lowry. 2010. "Evaluation of Target's Safe City Initiative: Implementing Public-Private Partnerships to Address Crime in Retail Settings." Washington, DC: The Urban Institute. http://www.urban.org/publications/412081.html.

Nichols, Austin. 2007. "Causal Inference with Observational Data." *Stata Journal* 7(4): 507–41.

Popkin, Susan J., Bruce Katz, Mary K. Cunningham, Karen D. Brown, Jeremy Gustafson, and Margery A. Turner. 2004. "A Decade of HOPE VI: Research Findings and Policy Challenges." Washington, DC: The Urban Institute. http://www.urban.org/publications/411002.html.

Popkin, Susan J., Molly M. Scott, Joe Parilla, Elsa Falkenburger, Marla McDaniel, and Shinwon Kyung. 2012. "Planning the Housing Opportunity and Services Together Demonstration: Challenges and Lessons Learned." Washington, DC: The Urban Institute. http://www.urban.org/publications/412516.html.

Popkin, Susan J., Brett Theodos, Liza Getsinger, and Joe Parilla. 2010. "An Overview of the Chicago Family Case Management Demonstration." Washington, DC: The Urban Institute. http://www.urban.org/publications/412254.html.

Robins, James M. 1997. "Causal Inference from Complex Longitudinal Data." In M. Berkans, Ed., *Latent variable modeling and applications to causality. Lecture notes in statistics,* 120: 69–117.

Robins, James M., Miguel Hernán, and Babette Brumback. 2000. "Marginal Structural Models and Causal Inference in Epidemiology." *Epidemiology* 11(5): 550–60.

Robins, James M., D. Blevins, G. Ritter, and M. Wulfsohn. 1992. "G-Estimation of the Effect of Prophylaxis Therapy for Pneumocystis Carinii Pneumonia on the Survival of AIDS Patients." *Epidemiology* 3:319–336.

Rubin, Donald B. 1986. "Statistics and Causal Inference: Comment: Which Ifs Have Causal Answers." *Journal of the American Statistical Association* 81(396): 961–62.

Smith, Robin E. 2011. "How to Evaluate Choice and Promise Neighborhoods." Washington, DC: The Urban Institute. http://www.urban.org/publications/412317.html.

Taubman, Sarah L., James M. Robins, Murray A. Mittleman, and Miguel Hernán. 2009. "Intervening on Risk Factors for Coronary Heart Disease: An Application of the Parametric G-Formula." *International Journal of Epidemiology* 38(6): 1599–1611.

Turner, Margery A., Austin Nichols, and Jennifer Comey. 2012. "Benefits of Living in High-Opportunity Neighborhoods: Insights from the Moving to Opportunity Demonstration." Washington, DC: The Urban Institute. http://www.urban.org/publications/412648.html.

Wilson, William Julius. 1987. *The Truly Disadvantaged.* Chicago: University of Chicago Press.

Witteman, Jacqueline C.M., Ralph B. D'Agostino, Theo Stijnen, William B. Kannel, Janet C. Cobb, Maria AJ de Ridder, Albert Hofman, and James M. Robins. 1998. "G-Estimation of Causal Effects: Isolated Systolic Hypertension and Cardiovascular Death in the Framingham Heart Study." *American Journal of Epidemiology* 148(4): 390–401.