Evaluating Spatial Policies[1]

Steve Gibbons (LSE, SERC, WWC) [2]

Max Nathan (NIESR, SERC, WWC) [3]

Henry G. Overman (LSE, SERC, WWC) [4]

Abstract

In most countries economic prosperity is very unevenly distributed. Regional, urban and neighbourhood policies are often based on concerns about these kinds of disparities, and reducing such disparities is a key policy objective in many countries. High quality evaluation is central to understanding how to meet these objectives. However, impact evaluation – which seeks to identify the causal effects of policies – is often in short supply for spatial policies. In this viewpoint we highlight three barriers that hamper more rigorous impact evaluation. First, data availability constrains research. Second, identifying the causal impact of polices is difficult. Third, there are several practical barriers. We briefly consider each of these in turn, and make practical recommendations for change. Better policy design, more use of open data, and capacity-building for government analysts are three important and achievable steps in improving the extent and quality of future impact evaluations.

Keywords: spatial policy; urban policy; evaluation

March 2014

[1] This viewpoint draws on a number of published sources: Gibbons, Nathan and Overman (2014); Gibbons McNally and Overman (2013); Gibbons and Overman (2012); Overman (2012)

[2] Steve Gibbons, Department of Geography and Environment, Spatial Economics Research Centre and What Works Centre for Local Economic Growth, London School of Economics, Houghton Street, London, WC2A 2AE, UK. Email: s.gibbons@lse.ac.uk

[3] Max Nathan, National Institute for Economic and Social Research; Spatial Economics Research Centre and What Works Centre for Local Economic Growth, London School of Economics, Houghton Street, London, WC2A 2AE, UK. Email: m.a.nathan@lse.ac.uk

[4] Henry G. Overman, Department of Geography and Environment, Spatial Economics Research Centre and What Works Centre for Local Economic Growth, London School of Economics, Houghton Street, London, WC2A 2AE, UK. Email: h.g.overman@lse.ac.uk

In most countries economic prosperity is very unevenly distributed. Regional, urban and neighbourhood policies are often based on concerns about these kinds of disparities, and reducing such disparities is a key policy objective in many countries. High quality evaluation is central to understanding if these objectives have been met.

There are different evaluation traditions. Process evaluation seeks to assess how public policies are designed and implemented, often using qualitative methods. Impact evaluation is instead concerned with identifying the causal effects of policy using quantitative methods (Gough et al, 2013). For many policies, the amount of process evaluation significantly outweighs that of impact evaluation. For example, in undertaking the first systematic review for the new What Works Centre for Local Economic Growth[5] we identified an initial list of around 1,000 articles that set out to provide evaluations of employment training policy. Of these studies, less than 100 provided a quantitative impact evaluation of the policy in question. Even then, many of these failed to score highly on the Scientific Maryland Scale that the Centre uses to rank the evaluation methodology and its implementation (Gibbons, McNally and Overman (2013).

Prior experience suggests that this example is likely to prove representative. The *impact evaluation* of many spatial policies falls short of the standards set by other policy areas – for example in clinical treatments, active labour market policy or international development. While this partly reflects inherent difficulties in spatial analysis, it also stems from a failure to adopt methods that could improve those evaluations.

In this *viewpoint* we highlight three barriers that hamper more rigorous impact evaluation. First, data availability constrains research. Second, identifying the causal impact of polices is difficult. Third, there are several practical barriers. We briefly consider each of these in turn.

The most common data problem arises from the lack of appropriate data at the appropriate spatial scale. For many spatial issues the correct unit of analysis is difficult to define, but administrative boundaries are likely to provide poor substitutes (Cheshire and Magrini, 2009). Sampled data creates additional problems, particularly at smaller spatial scales. For a given sampling rate, smaller spatial scales reduce the average sample size for each spatial unit. For these reasons, there can be substantial problems generating descriptive statistics, even for administrative units.

Even though administrative units may be arbitrary from an analytical perspective, they are, of course, very important to policy makers: local leaders and officials are always keen to know how their places are 'performing'. As a result, local policymakers sometimes worry that without comprehensive data for (say) a Local Authority district, there is no way they can understand the impact of spatial policies for that local authority.

[5] www.whatworksgrowth.org

Happily, this need not be the case. Sampled data *can* be informative about the impact of policy, even if it cannot comprehensively describe outcomes for specific places. For example, a small sample of firms from a local authority may not give a precise estimate of average firm size in that local authority. However, if a policy aimed at increasing firm employment is applied across several local authorities, then combining samples from different local authorities may still provide estimates of the impact of the policy.

This might raise concerns for those who believe that every location is unique in every aspect. However, most empirical analysis – qualitative and quantitative – proceeds under the assumption that the cases under investigation, including the responses to policy changes, share at least some characteristics. This allows researchers to construct counterfactuals, predictions of what would have happened in the absence of policy (on which more below). Impact evaluation then needs to take account of the heterogeneity between places when formulating research designs and interpreting results – but this does not rule out all evaluation *per se*. In short, while the lack of large samples for administrative spatial units may appear to be a barrier to evaluation, it need not be.

Even when appropriate spatial data is available, many impact evaluations have not paid sufficient consideration to the crucial issue of identification of causal effects. These problems are discussed in detail in Gibbons and Overman (2012). This neglect of causality has profound implications for our understanding of policy effects.

Causality is concerned with questions of the type 'if we change x what do we expect to happen to y?' The fundamental challenge to answering these questions is that policies are not usually randomly assigned. As a result, in most real-world cases we lack the *counterfactual* that tells us what would have happened to the targets or recipients of the policy ('the treated') if they had not been treated: this is fundamentally unobservable. This is a problem, because it is the comparison of actual outcomes to this counterfactual that identifies the causal impact of policy. So the way in which this counterfactual is (re)constructed is the key element of programme evaluation design. Applied economics has come a long way in its efforts to find credible ways to construct such counterfactuals from observed data (Angrist and Pischke, 2009). Unfortunately, however, such methods have not been widely used in the evaluation of spatial policy.

A standard approach is to create a counterfactual group of similar (say) individuals not participating in the programme being evaluated. Outcomes can then be compared between the 'treatment group' (those affected by the policy) and the 'control group' (similar individuals not exposed to the policy). The challenge for good programme evaluation is to ensure and demonstrate that this control group is plausible.

If the construction of plausible counterfactuals is central to good policy evaluation, then the crucial question becomes: can we design such counterfactuals for spatial policies? The answer is certainly ‘yes’ for some, but not all, spatial policies. The answer may depend crucially on the extent to which evaluation was embedded in the policy design process so that the policy was implemented in such a way as to allow the construction of a counterfactual.

One way to achieve this is to randomly assign cases to treatment and control groups. Such Randomised Control Trials (RCTs) are often considered the ‘gold standard’ of evaluation (Banerjee and Duflo 2010, DiNardo and Lee, 2010, Haynes et al 2012). Properly implemented, randomisation ensures that treatment and control groups are comparable, thus identifying the causal impact of policy (Katz et al 2001, Kling et al 2005, Sanbonmatsu et al 2012). However, there are challenges applying RCTs to spatial policies. For example, clinical trials are typically halted when significant evidence of benefits / harm first emerges. This can be harder to implement for spatial policies, where effects may take months or years to appear, and where effects may be harder to reverse. Large-scale experiments are costly and can still suffer from design flaws; small scale experiments may not generalise to other contexts.

Where randomised control trials are not an option, there are various ‘identification strategies’ that can be used instead. For area-based interventions, one possible approach is to compare those areas treated to other similar areas that were not treated. Such simple comparisons remain problematic, however: unless we have an exhaustive list of area characteristics that influence outcomes, we should worry that some unobserved characteristic drives both the decision to target the area *and* outcomes in that area (Criscuolo et al 2012). In this case, we would wrongly attribute changes in outcomes to the policy when, in fact, they are driven by unobservable area characteristics. Much of the improvement in policy evaluations has come from novel ways of addressing this problem, combined with better understanding of how to interpret the results.

One possibility is to compare outcomes for areas that receive funding to areas that applied for, but did not receive, funding. Busso et al (2013) apply this approach to US Empowerment Zones, and it could also be used for competition-based programmes in the UK. We could extend this idea to a three-way comparison between these two groups and a group of areas that did not apply for funding. The timing of policy interventions may provide another source of identification: in theory, if some areas are given money before others they should start improving earlier. If not, that raises questions about whether treatment caused any improvement.

Even if we cannot be sure that funding decisions are uncorrelated with *all* relevant unobservable characteristics, we may believe that this condition holds for *marginal* decisions. For example, government may make funding decisions on the basis of a ranking of projects or of areas. If this ranking is available, then this may allow identification from the comparison of outcomes for areas just ‘above the bar’ (and treated) to outcomes for similar areas just ‘below the bar’ (and not treated). In other cases, changes to policy design may allow identification. For example, if firms are only treated in eligible areas, then changes to the map of eligible areas may allow identification by comparing treated firms to similar firms who might have been treated before the map changed, but are now ineligible. Criscuolo et al (2012) use such shifts in EU eligibility criteria to identify the effects of Regional Selective Assistance on UK firms.

The fundamental idea underlying all these approaches is that, in the absence of explicit randomisation, ‘quasi-experimental’ sources of randomisation may address selection on unobservables. These sources may occur as a result of institutional rules and processes (and changes in these), or through environmental or other phenomena that result in some cases randomly receiving treatment.

Even using these strategies, though, the treatment and control groups may not be comparable. Statistical techniques such as Ordinary Least Square (OLS) and matching can be used to address this problem. However, good quality impact evaluation uses identification strategies to construct a control group and *then* tries to control for remaining differences on observable characteristics. It is the combination that is powerful: OLS and matching alone raise concerns about the extent to which unobservable characteristics determine both treatment and outcomes and thus bias the evaluation.

Evaluations of spatial policies paid for by government usually make little use of these strategies. This significantly complicates policy development, because reports that are less careful about causality often make much broader claims about the impact of policy (and how that impact was achieved). As a result, policy makers face a difficult trade-off when trying to decide how to evaluate policies. Wide-ranging ‘evaluations’ that are less careful about causality *appear* to provide more information. In contrast, high-quality quantitative evaluations often make fairly narrow claims about the impact of policy.

Improving evaluation need not be costly. Using policy design to assess causal effects ideally requires detailed information about the decision making process. How were bids solicited and assessed? How were the winning bids selected? How were funding levels decided? Unfortunately, in many cases crucial information (e.g. for unsuccessful applicants) is either not systematically recorded or is not made available to researchers. For all kinds of reasons, governments remain reluctant to change this situation.

A second, trickier challenge is officials’ and elected leaders’ (understandable) desire for rapid results. High quality evaluation requires data on both policy and relevant outcomes. That outcome data is usually only available with a time lag – which complicates the interaction between evaluation and policy formulation, if policy makers are working on shorter time scales. In short, better evaluation often requires patience and transparency. Unfortunately, political imperatives – in particular, the desire to show quick results – can over-ride the need to take a longer-term view.

Some of these problems stem from fundamental differences in goals between researchers and policy makers. Happily, others are more easily addressed. Collecting data for more 'sensible' spatial units – such as metropolitan areas – can better align the spatial scales used by policy makers and analysts. Designing policies with clear, well-documented rules and easily observable recipient characteristics helps researchers use programme features to evaluate those policies. Using such institutional features of policies will, in turn, help improve impact evaluation. Other important steps are opening up programme datasets to researchers, and improving government analysts' capacity to understand and work with the newest impact evaluation techniques.

Of course, simply having (process or impact) evidence to hand is not sufficient to make good policy: policymakers need objectives, and principles to guide these. But belief or principle-driven policy making is also often costly and ineffective, and such approaches will still win out over evidence-based policy making in many situations. Strong evidence is a necessary condition for effective policy making, and addressing the barriers above will help move us towards this. Addressing these problems also makes for better research and evaluation – regardless of any influence on policy.

References

Angrist, J. and J.S. Pischke (2009). Mostly Harmless Econometrics. Princeton, NJ: Princeton University Press.

Banerjee, A. and E. Duflo (2009). The Experimental approach to development economics. Annual Review of Economics 1: 151-178.

Busso, M., J. Gregory, and P. Kline (2013). Assessing the Incidence and Efficiency of a Prominent Place Based Policy. American Economic Review, 103(2): 897-947.

Cheshire, P. and S. Magrini (2009). Urban growth drivers in a Europe of sticky people and implicit boundaries. Journal of Economic Geography 9: 85-115.

Criscuolo, C., R. Martin, H. G. Overman and J. van Reenen (2012). The causal effects of an industrial policy. Spatial Economics Research Centre Discussion Paper SERCDP0098. London, SERC.

DiNardo, J., and D. S. Lee (2010). Program evaluation and research designs in Ashenfelter, O. and D. Card (eds) Handbook of Labor Economics Vol 4. Elsevier.

Gibbons, S., M. A. Nathan and H. G. Overman (2014). Evaluating Spatial Policies, SERC Policy Paper No. 12.

Gibbons, S., S. McNally and H. G. Overman (2013). Review of government evaluations: A report for the UK National Audit Office. http://www.nao.org.uk/wp-content/uploads/2013/12/LSE-Review-of-selection-of-evaluations-with-appendices1.pdf.

Gibbons, S. and H. G. Overman (2012). Mostly pointless spatial econometrics. Journal of Regional Science, 52, 2, pages 172–191.

Gough, D., S. Oliver and J. Thomas (2013): Learning from Research: Systematic reviews for informing policy decisions, London: Alliance for Useful Evidence.

Haynes, L., O. Service, B. Goldacre and D Torgerson (2012). Test, Learn, Adapt: Developing public policy with randomised control trials. London, Cabinet Office.

Katz, L. F., J. R. Kling and J. Liebmann (2001). Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment. The Quarterly Journal of Economics 116(2): 607-654.

Kling, J. R., J. Ludwig and L. Katz (2005). Neighborhood Effects on Crime for Female and Male Youth: Evidence from a Randomized Housing Voucher Experiment. The Quarterly Journal of Economics 120(1): 87-130.

Overman, H. G. (2012). Geographical Economics and Policy. In Fischer M M,Nijkamp P (Eds.) Handbook of Regional Science. Springer Heidelberg New York Dordrecht London

Sanbonmatsu L, Marvakov J, Potter N, Yang F, Adam E, Congdon WJ, Duncan G, Gennetian L, Katz LF, Kling JR, Kessler R, Lindau ST, Ludwig J, McDade T (2012). "The Long-Term Effects of Moving to Opportunity on Adult Health and Economic Self-Sufficiency." Cityscape 14(2): 109-136.