Place-based policies for the future

Things We Don't Want to Know? Monitoring and evaluating place-based policies

Max Nathan, UCL and CEP

Paper presented at Workshop 4: "Measuring progress better: Cutting-edge approaches to monitoring and evaluating place-based policies", 7 July 2023, Paris

Good monitoring and policy evaluation is important to effective placed-based policymaking, especially now. But robust evaluation evidence is hard to generate. It is also surprisingly hard to incorporate into policymaking infrastructures. This paper will consider why this is, and will suggest ways forward. I first discuss why we care about monitoring and evaluating place-based policies. Then, I outline the main questions for their evaluators. I highlight the problem of incomplete evaluation evidence, especially in impact evaluations (which start with the question 'what's the effect of policy X on outcome Y?') I then review how to build an impact evaluation strategy for place-based programmes. I focus on the core design issue - establishing a counterfactual - then methods and data. I use EU Cohesion Policies and broadband support programmes as case studies. I discuss what we know about the effectiveness of these interventions. Next, I set out challenges to generating and using evidence. These tasks combine practical challenges (things that are hard to know) and cultural / institutional challenges (things we don't want to know). Finally, I discuss what we can do to tackle these challenges. I draw on UK experience from the 'What Works' movement over the last 10 years.

Funded by
the European Union

ABOUT THE OECD

The OECD is a multi-disciplinary inter-governmental organisation with member countries which engages in its work an increasing number of non-members from all regions of the world. The Organisation's core mission today is to help governments work together towards a stronger, cleaner, fairer global economy. Through its network of specialised committees and working groups, the OECD provides a setting where governments compare policy experiences, seek answers to common problems, identify good practice, and co-ordinate domestic and international policies.

More information available: www.oecd.org. OECD Centre for Entrepreneurship, SMEs, Regions and Cities on Twitter: @OECD_local

DISCLAMERS

This paper was prepared as a background document for an OECD-EC High-Level Workshop on "Place-Based Policy for the Future". It set the basis for reflection and discussion. The background paper should not be reported as representing the official views of the European Commission, the OECD or one of its member countries. The opinions expressed and arguments employed are those of the author(s).

This publication was produced with the financial support of the European Union. Its contents are the sole responsibility of the author and do not necessarily reflect the views of the European Union.

This document, as well as any statistical data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

ABOUT THE OECD-EC HIGH-LEVEL WORKSHOPS ON PLACE BASED POLICIES FOR THE FUTURE

The series of six OECD-EC High-Level Workshops on "Place-Based Policies for the Future" consider place-based policies from the following angles: (1) how place-based policies have evolved; (2) the articulation of place-based policies against other policymaking frameworks; (3) how place-based policies can address current and future societal problems; (4) policy design, measurement and evaluation; (5) governance and coordination; and, (6) what place-based policies are required for the future.

The outputs of the workshops are a series of papers and a summary report that outlines future directions for place-based policies. This work will ultimately be relevant for policymakers at all levels of government who are interested in improving the design and implementation of place-based policies to contribute to equitable and sustainable economic futures.

The workshops support the work of the OECD Regional Development Policy Committee and its mandate to promote the design and implementation of policies that are adapted to the relevant territorial scales or geographies. The seminars also support the work of the Directorate-General for Regional and Urban Policy (DG REGIO) of the European Commission. The financial contributions and support from DG REGIO are gratefully acknowledged.

Acknowledgements

Table of Contents

Figures

Overview

Good monitoring and policy evaluation is extremely important to effective placed-based policymaking, especially now, when there is a great appetite for forward-looking place-based interventions. However, robust evaluation evidence is both hard to generate, and surprisingly hard to incorporate into mainstream policymaking infrastructures. This paper will consider why this might be and will suggest some ways forward – for researchers and for policymakers.

This paper first discusses why we should care about monitoring and evaluation of place-based policies. I start by defining place-based policies. I then set out why monitoring and evaluation of place-based policy is especially salient now. I highlight increased awareness of persistent area disparities, shifting views about the effectiveness of place-based interventions, and the current hot topics of big-state industrial policy and industrial missions in the EU, US and UK, all of which involve significant place-based components.

I go on to discuss the main kinds of evaluation questions that evaluators of place-based policies should be asking, show how these questions are connected, and discuss the overall state of the evidence, and the key problem of missing or incomplete evaluation evidence. This problem is particularly germane for ex-post evaluations (as opposed to ex-ante studies), and for impact evaluations (as opposed to process evaluations).

Second, the paper briefly reviews the core components of an evaluation strategy for place-based programmes. I focus on impact evaluations (which starts with the question 'what is the overall effect of policy X on outcome Y?'). I use EU Cohesion Policies as an

example of a classical place-based policy, and broadband support programmes as an example of place-sensitive policy. I cover the core design issues, focusing on establishing a counterfactual, then move to the typical toolkit of methods and data types evaluators bring to these cases. I also look at what we know about the overall effectiveness of these interventions.

Third, the paper sets out the challenges to generating and using evidence, both to implement monitoring and evaluation of place-based policies, and to mainstream the findings into political space and into decision-making institutions. In essence, monitoring and evaluation involves both practical challenges – things that are *hard to know* – but also cultural / institutional challenges, things *we don't want to know*.

Practical challenges include establishing causal effects, which is especially hard in area-based settings; grappling with variable data access and quality, long impact timeframes, varying government capacity across scales and countries, and assessing non-economic outcomes which evaluation methods find it hard to proxy at scale. Other challenges are less tangible but equally, if not more important. We – academics, policymakers – spend a lot of time thinking and talking about place-based policy. And yet we often lack evidence and make it hard to gather evidence on its effectiveness; and incorporating lessons often feels harder than it should be. I will conclude by setting out some high-level ideas about what we might do to tackle these challenges, drawing largely on UK experience from the 'What Works' movement over the last 10 years.

1 Setting the scene: why do we need to evaluate place-based policies?

1.2 / Defining place-based policies

'Place-based policy' is a broad field, and one which is evolving over time (Cheshire, Nathan et al. 2014, Neumark and Simpson 2015). Defining the policy space is therefore not straightforward. Here I follow the very useful discussion by Suedekum (2021 , 2023) who distinguishes four main features of the space.

First, we can pick out two main flavours of policy: direct and indirect. *Direct interventions* target specific places; or people/firms in those specific places, and aim to help lagging places catch up to leading places. This is what most of us would understand by "place-based policy". Examples would include US and European Employment Zones, US Empowerment Zones, Freeports, and at larger scale, EU Objective 1 and 2 programmes in the 1980s-2000s. More recently, there has been a shift in the EU and other countries towards *indirect interventions*. These "place-sensitive policies" are aimed at many kinds of places, both leaders and laggards, and have a mix of objectives: sometimes catch-up, sometimes helping consolidate particular forms of area comparative advantage (like clusters) or helping areas transition from one development path to another. Examples would include EU Cohesion Policy from 2014 to present, the emerging UK Strength in Places Fund and City / Growth Deals, and other programmes to improve areas' physical and digital infrastructure. Examples of the former would include providing new road, train or light rail lines, or via service improvements to existing networks. Examples of the latter would be through supported provision of broadband/satellite/mesh networks, and/or vouchers to raise uptake / adoption of existing services.

As Suedekum points out, programmes also differ in their funding sources and structures: these can be dedicated / targeted (vertical) or mainstream (horizontal). The former involves funds spent directly on a place from national (or European) levels of government; the latter involves transfers from one place to another, via national government. [1] In any given OECD member state, the balance of place-based funding varies according to membership (or otherwise) of supra-national institutions, national institutional and policy design choices, and crucially the level of devolution within that member state.

In many countries, place-based policy tools are distributed across central, regional and local government scales and budgets, making it harder to 'see' – and evaluate – as a coherent set of functions. To take a possibly extreme example, in the UK local economic development lives simultaneously in local government, planning, business, policing, employment, transport and housing functions of central and local government, as well as in the Treasury. Conversely, UK places' local tax base is very limited (around 70% of local govt spending comes from central govt (CFC REF)), so horizontal reallocation mechanisms are very important. [2]

We can see how these institutional features also matter to monitoring and evaluation in a number of ways, notably through influencing a) overall policy coherence b) and relative power / capacity of actors in the 'evaluation ecosystem'. I will return to these issues further in Part 5 of the paper.

---

[1] Almost all public policies have *some* spatial dimension, either explicit or implicit. Here I focus on programmes where the place element is a) explicit and b) a major feature of the programme.

[2] A concrete example is business property tax, which until very recently, involved rates set centrally, collected locally, then redistributed by the centre as part of block grant to local areas.

1.2 / Place-based policies in the current moment

A useful way into understanding the importance of effectively monitoring and evaluating place-based policies is to consider their current policy salience. Place-based policy really *matters* right now, in other words, and so properly evaluating its effects is essential. There is currently a great appetite for forward-looking place-based interventions. There are at least three big reasons for this.

The first is that in most OECD member states, there are long-running spatial disparities in income, wages, employment, health and life expectancy between cities and regions: see Brandily et al (2022) for UK evidence, Von Ehrlich and Overman (2020) for EU evidence, and Overman and Xu (2022) and OECD (2022) for international evidence. Spatial economic disparities are always hard to shift, not least because they are partly self-reinforcing: agglomeration economies in leading places increase more qualified workers' productivity and nominal wages; in turn this leads more knowledge-intensive firms and skilled workers to sort into those places, widening gaps with other places. At the same time, these forces also generate important *within-place* inequalities between different worker groups: notably, in US cities – and likely elsewhere – the urban wage premium that raised all wages in big cities has been breaking down for non-graduates, and especially for the least qualified (and lowest paid) members of the workforce (Autor 2019).[3] The structural mega-forces driving these trends seem very challenging to shift (Cheshire, Nathan et al. 2014). Nevertheless, over recent decades there are also important examples of physical and economic transformation in places' fortunes, both at individual city scale and wider regional scale, and it is also clear that public place-based policies have played a non-trivial role in this (What Works Centre for Local Economic Growth 2016, Enenkel 2021). Robust monitoring and evaluation of place-based policy is

---

[3] Relatedly, Diamond and Moretti (2021) look at differences in cost of living across US commuting zones.

essential to understanding policy effects, the underlying mechanisms and their reproducibility.[4]

Second, as Suedekum (2023) and McCann (2023) both note, there has been an important shift in academic conventional wisdom about place-based policy in the past decade. This change in views has led some previously sceptical policymakers to take place-based programmes more seriously; in the US and EU, at least, the effects of this view change can be seen in emerging policy frameworks. These shifts make effective monitoring and evaluation of those programmes more important.

In the UK, for example, place-based policy typically had strong advocates in the local government and central government 'place' functions (now the Department for Levelling Up, Housing and Communities)[5] and the local economic development research community. Conversely, the Treasury (department of finance), business department, and most academic economists were deeply sceptical about place-based initiatives. The primary of the Treasury (Coyle 2021) ensured that place-based policies remained relatively marginal and small-scale. Economists typically recognised the strong equity case for place-based initiatives, but argued that in a spatial economy with mobile firms and workers, the efficiency case for place policy was generally weak. Rather, policy should support people and firms directly (Glaeser and Gottlieb 2008). The subsequent shift in economists' and economic policymakers' views of place-based policy has at least three causes. First, even in a world of people/firm-based policy, place-based *delivery* may be an effective way to target resources on key groups and individuals (Tunstall and Lupton 2003). Second, as Suedekum (ibid) explains, the broader efficiency case for place-based policy is now stronger. This is partly thanks to new theoretical models which emphasise that real-world spatial economies can be characterised by both negative

---

[4] In this context, it is particularly important for evaluators to distinguish between programme effects on the target region or place, and the relative effects on other places. Kline and Moretti (2014) is a good example of an evaluation that does both, for a New Deal-era regional programme in the US.

[5] https://www.gov.uk/government/organisations/department-for-levelling-up-housing-and-communities

spillovers from congestion, and positive spillovers from clustering (Kline and Moretti 2013, Fajgelbaum and Gaubert 2020, Moretti 2022). It also derives from new evidence which emphasises the limited mobility of people, for example in the UK (Bosquet and Overman 2019) and the US (Molloy, Smith et al. 2017).[6] Third, there is a broader political economy imperative to respond to populist movements across EU and OECD member states, which have at least some roots in economic and political grievances about area-based disparities.[7]

A third set of factors driving the resurgence of place-based interventions is the 'polycrisis' of the pandemic, climate change and geopolitical instability, alongside localised populist movements (Tooze 2022). Faced with these risks, leaders in many OECD member states are rolling out, or contemplating very large-scale economic interventions – with strong place-based components, such as those discussed in the next paragraph (Bailey, Pitelis et al. 2023). We can see this immediate policy moment as part of a larger, longer-term resurgence in industrial strategy (Rodrik 2004, Tirole 2017), especially the current emphasis on mission-orientated policy (Mazzucato 2018). While industrial policy of the past has not always had a clear place-based element, these new / emerging strategies typically combine a mix of regulatory, sectoral and technology-focused interventions (to secure supply chains, enable a climate transition, and green the economy) with place-based and place-sensitive elements (to support the growth of co-located activities, rebuild a national industrial base, and to enable lagging places to benefit from structural shifts).

Most notably, in the US, the Inflation Reduction Act, CHIPs Act and Infrastructure Bills have so far resulted in a surge of investment in high-tech manufacturing across the US,

---

[6] Together, these create a more powerful case for place-sensitive policies to boost agglomeration, and to redistribute activity across places; and start to undercut the arguments that targeting places inevitably creates deadweight loss.

[7] Geographies of discontent do not always map onto geographies of deprivation. For example, the Brexit vote partly reflects the 'revenge of places that don't matter' (Rodríguez-Pose, 2018), but is more accurately described as a coalition of older non-graduate voters, in both deprived and affluent locations (Green and de Geus, 2022).

especially in historically cheaper and rust-belt locations (Sullivan 2023). At the same time, the Tech Hubs initiative aims to develop nascent clusters of such activity, using Triple-Helix-type strategies (Muro, Parilla et al. 2023). At the time of writing, the EU is about to unveil a draft 'Green Deal Industrial Plan', which is likely to include a European Sovereignty Fund for cleantech and some looser limits on state aid, on top of the Net-Zero Industry Act and Critical Raw Materials Act (Crawford 2023). The UK's story has many more twists and turns, with repeated changes to the policy mix, frameworks and institutions (Nathan 2022, McCann, Ortega-Argilés et al. 2023). Nevertheless, the opposition Labour Party has developed five national missions, including an economic mission underpinned by a green capital investment plan of up to £28bn / year, including an emphasis on building high-value green industries. The explicit aim is to use this big push to spread economic opportunity across the country and to tackle spatial economic disparities (Reeves 2023).

Again, these policy dynamics also raise the importance of good monitoring and evaluation. The strategies outlined above involve extensive, well-resourced programmes which are being rapidly developed and deployed. There is much less clarity about exactly how the fine grain of programmes will work, and what we can expect these to do (Tooze 2023). Building in robust systems to track and test programme impact will be essential to ensuring their legitimacy, and crucially, maximising their longer-term effects.

2 What do we need to know?

Part 1 of this paper made the case for effective monitoring and evaluation of place-based policies. This part summarises the key questions researchers and policymakers should be asking. The main dimensions are set out in Figure 1, below.

Figure 1. Types of project evaluation.
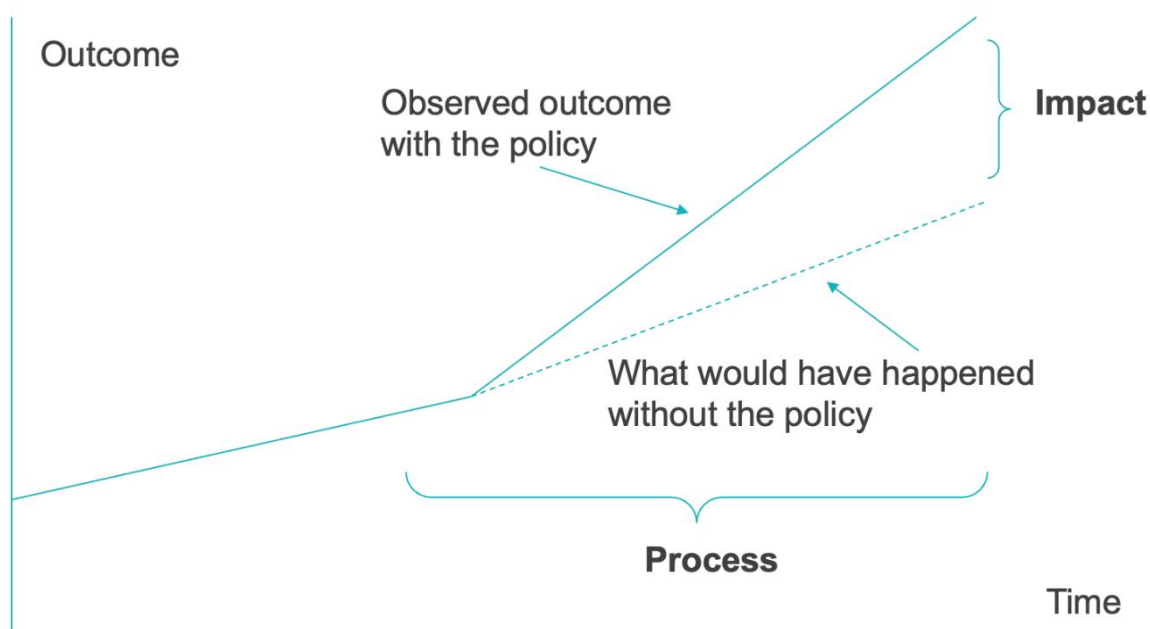
The starting point in any policy evaluation is a theory of change, or 'logic model': this sets out, usually graphically, how we expect a policy to achieve its objectives: from inputs, to actions, to outputs, to outcomes (Bates and Glennerster 2017, What Works Centre for Local Economic Growth 2022). The logic model informs programme monitoring, by specifying a range of input, activity and output data to gather. The logic model also helps us understand both how to frame specific evaluation questions (what outcomes to test), and how to answer them (what data to use, and what channels to investigate).

In principle, evaluators need to ask multiple questions of a policy (HM Treasury 2020). When developing a policy, we might first want to model the possible effects of a policy: that is, what might be the effect of the policy on the desired outcomes. Evaluators refer to this as *ex-ante modelling*: it is closely related to building the logic model / theory of change. After this, the main questions are about the actual effect of the policy on those outcomes. Evaluators refer to this as *ex-post impact evaluation.* Typically, doing an impact evaluation involves asking one or more of the following:

- What was the overall effect of policy A on outcome Y?
    - What explains this effect?
- What has a bigger effect on outcome Y, policy A or policy B?
    - What explains this?

The second order questions we need to ask are about *effect heterogeneity* and about *process (i.e. "process evaluation")*. The first of these is concerned with the effects of policy in different places, or on different groups of policy and people (and in particular, whether some groups do better or worse than others). The second of these, as the name suggests, is concerned with the process of implementing a policy, and what worked more or less well in the rollout. Addressing these concerns will typically involve asking things like:

- What works (or what works best) where?
    - What e.g. contextual / institutional factors explains this?
- Who does the policy work for? Who are the winners and losers?
    - What explains this?
- What is the most effective way to implement the policy?

       o   How can we replicate and/or scale it?

These questions may all seem straightforward enough. However, they come with three immediate caveats which make them more challenging.

First, all these questions are complements, not substitutes. For example, process evaluation a valuable complement to impact evaluation – as many readers will already recognise. Similarly, understanding the winners and losers from a policy, and whether it will work 'in my place' are central questions for practitioners. As researchers and as policymakers, we should want to answer all of them, for a given policy.

Second, though, for any given policy it is very hard to answer questions about process, heterogeneity and distributional impacts without answering the first-order questions about overall impact. There is a clear, 'natural' sequence here.

Third, for many policies, robust evaluation evidence is often incomplete, or missing – and this is particularly the case for impact evaluation evidence. In the UK, for example, Sanders and Breckon (2023) report two National Audit Office exercises that found between 2006 and 2012, only half of central government departments had counterfactual-based impact evaluations of policies; and of 108 recent major strategic projects, costing £432bn in 2020 money, only nine had been impact-evaluated robustly; 77/108 had *no evaluation arrangements at all*. As Sanders and Breckon put it:

> *"Almost nobody is against evidence-informed policy and practice. Like motherhood and apple pie, there's not a lot there to object to. After all, who wants to take to a stage and argue in favour of evidence-free policy? Yet, despite this near universal support, the use of evidence has been far from inexorably rising. Zombie policies – those that we believe had been killed off by evidence – rise again and again, while the truly evidence-based struggle to gain traction."* [p3]

So why might this be? One way into this is through Figure 2, below, which sets out some core best practices for evaluating local economic development policies, as developed by the UK's What Works Centre for Local Economic Growth. It is designed to cover policies aimed at people and firms, as well as places, so fits within the scope of place-based policies outlined earlier.

Figure 2. Best practices for project evaluation.

START EARLY
Evaluation is cheaper and more effective if it is intergrated into project design

DEFINE SUCCESS
Employment and productivity are good measures of local economic growth

WHAT TO EVALUATE?
Do you want to focus on overall effects or on what works better

FIND A CONTROL GROUP
Demonstrate your project's impact by comparing participants with similar groups that have not been involved

COLLECT DATA
Keep track of who participates and what happened to them before and after participating

HOW LONG?
Short evaluations won't capture all the effects, but longer ones are more expensive

PLAGIARISE!
Use previous evaluations as templates

GET EVERYONE ONBOARD
Give delivery partners clear instructions about what type of evaluation is acceptable

Source: What Works Centre for Local Economic Growth (2022)

These practices probably feel intuitive. So why might they not happen? I would suggest that the Figure actually shows us a mix of actions that are a) likely to involve technically challenging tasks (defining research questions, finding a control group, collecting data, measuring outcomes over multiple timeframes) and b) likely to involve things that are practically challenging (building in evaluation design and monitoring systems at the outset, setting clear policy objectives, building partnerships and alliances to help implementation). These latter are especially challenging under limited institutional resources, autonomy or capacity. We will return to both sets of challenges in the rest of the paper.

3 Designing impact evaluations

Parts 1 and 2 of this paper set out the case for robust evaluation of place-based policies, the key questions evaluators need to answer, and the problem of incomplete evidence. This part gives a technical overview of the evaluation process, using two examples to drive the discussion: EU Cohesion Policy, a classical place-based policy, and tools for improving digital infrastructure, which typically involves place-sensitive approaches. For reasons of space, I focus on impact evaluation issues.

3.1 / Tools for understanding impact

This section draws heavily on evidence reviews by the UK's What Works Centre for Local Economic Growth (2015, 2016).

For any place-based policy, once the theory of change is in place, the first order evaluation questions are about the *overall impact* of the policy on the desired outcome or outcomes. What is the *effect* of that policy on the outcome, or, what is the *most effective variant of the policy* on that outcome? To be confident about true effects we need to have some confidence our impact evaluation is picking up the *causal effects* of the policy on the outcome. In the main tradition of impact evaluation, the central challenge in 'identifying' causal effects involves building a valid counterfactual: what would have happened to a place (or the people / firms in it) if the programme hadn't happened. Without a time machine, we can't know this, so we need to construct the counterfactual using statistical / econometric techniques.[8]

---

[8] See Cartwright (2009) for a discussion of alternative traditions of establishing causal chains. Some of these, including detailed case studies / case work, are valuable complements to the toolkit discussed here.

The workhorse approach in place-based policy impact evaluation is to build a counterfactual group of places similar to those where the policy is happening. We can then compare changes in outcomes in places where the policy happened (the 'treatment group') to changes in similar places where the policy did not happen (the 'control group'). The same logic applies to policies targeted on people or firms in specific locations.

The key question for impact evaluation, then, is how robustly we can build this counterfactual. As shown below, there are several good ways to do this for place-based policies, depending on the policy design, implementation setting and availability of data. It is important to be clear that for place-based policies of whatever flavour, building the counterfactual is challenging. It is something that is *hard to know*.

One key issue is what economists call 'selection into treatment', which is when places in a policy differ from those who do not. For example, EU Cohesion Policy typically targets less developed regions, which have weaker economies than richer regions, and are on different growth trajectories. If we don't control for this, we risk confusing features of the treated places with effects of the policy. Specifically, even if Cohesion Policy accelerates growth in poorer places, richer places may still grow faster because of their underlying growth trajectory. So, if we compare changes in growth for poorer and richer places, we are not taking into account the underlying differences between the two groups of places. Our estimates of policy effects will be an under-estimate of their true effect.

For programmes aimed at leading regions, we have the opposite problem: upwards bias on the true effect. For example, broadband rollout programmes may involve both of these issues, depending on the policy objective: some are targeted at areas poorly-served by existing internet providers, others target high-performing places where poor infrastructure is deemed to be a constraint on economic growth.

If we can see how treated and control areas differ, and proxy for this in our data or design, then we can control for these differences in our analysis – making it easier to isolate the policy effect we are interested in. But in many cases, we have the second key issue that selection happens in ways that are either hard to proxy – or hard to observe at all. And third, underlying differences between treated and control areas are often likely to vary over time, complicating the picture further: even if we can control for selection into the programme at the outset, we will need statistical tools to handle unobservables while the programme is running, and that may influence outcomes.

Dealing with these technical challenges typically requires evaluators to leverage some element of randomness in the programme. Crucially, for policies focused on places, randomised control trials (RCTs) – the 'gold standard' of impact evaluation (Heffetz and List 2021) – are almost always out of scope. Fundamentally, randomisation is very challenging for capital investments and for programmes that target places. Such 'treatments' are usually expensive, often take time to take effect, and are not easily reversible. This creates serious technical and ethical issues, so for the most part RCT designs are not used in this space.[9] Nevertheless, RCTs are feasible for place-sensitive policies that target firms or people, and some of the ethical concerns can be mitigated by designs which either offer varieties of a policy, or that randomise the timing of policy. Both approaches help ensure that no-one is left untreated. Two UK examples are the Creative Credits programme in Greater Manchester, which randomised vouchers for SMEs in the city to procure from firms in the local creative industries cluster (Bakhshi, Edwards et al. 2013), and the national Growth Vouchers programme, which randomised vouchers for SMEs to purchase strategic business advice / consultancy services.[10]

---

[9] RCTs have been used in some US area-based programmes, notably interventions like Moving To Opportunity, in which involve demolishing / rebuilding social housing estates, and relocating existing residents.

[10] https://www.gov.uk/government/collections/growth-vouchers-programme, accessed 10 August 2023.

Box 1: impact evaluation designs for EU Cohesion Policies and broadband policies

Several **EU Structural Fund programmes** have/had similar differences in eligibility rules, and these can be used to generate estimates of programme impact. It is plausible that selection into these programmes is based on unobservable as well as observable factors. Evaluators can then generate treatment and control groups of regions which are observably similar in their economic conditions and trajectories, and which only differ in their eligibility into the programme over time. Comparing areas either side of the cutoff – a regression discontinuity design – helps achieve reliable estimates of the programme impact. Examples of this approach include Becker et al (2010, 2013) and Pellegrini et al (2013), all for Objective 1 funding. Criscuolo et al (2019) also exploit changes in eligibility rules when assessing the UK's Regional Selective Assistance programme. By contrast, Becker et al (2012) focus on the effects of different *levels* of Objective 1 support. For the set of supported NUTS3 regions, they use propensity score matching to identify places with similar characteristics but different levels of funding 'intensity'. Comparing changes over time allows them to back out the effect of different levels of Objective 1 intervention.

Evaluators worry that ISPs typically offer **broadband access** first to places where demand is highest. If not correct for, this biases up the effects of rollout. A paper by Akerman et al (2015) uses features of a national policy to get at the causal effects of broadband rollout in Norway. Policymakers designed a scheme to roll out broadband nationally; however, cost constraints meant that firms in some areas got the technology sooner than others. Because the cost factors were related to the policy and not market forces, this allowed the researchers to 'back out' the impact of broadband on recipient versus non-recipient firms. More technically, several papers evaluating early broadband rollout programme impact have used properties' distance to the nearest telephone exchange as a proxy for broadband access (Ahlfeldt et al 2017 (2017), Sanchis-Guarner et al 2021 (2021). Initial rollout used existing phone networks, and properties beyond a certain distance to a local exchange could not be connected. This created a form of randomness between otherwise similar properties, firms and people either side of the cutoff distance. In turn, this could be used as an unbiased proxy for 'access to broadband', either using an instrumental variables or regression discontinuity design.

Drawn from What Works Centre for Local Economic Growth (2015, 2016).

If RCTs are off the table, the next best option is to exploit 'quasi-randomness' in the programme. While this sounds obscure, in practice it usually involves exploiting quite familiar features of place-based programme design and implementation. Indeed, one reason why impact and process evaluations are highly complementary is that it is often impossible to design a robust impact evaluation of a given place-based without detailed knowledge of that policy's origins, actors, internal debates and on-the-ground rollout. What does this approach look like in practice? Box 1, above, gives some examples.

3.2 / Methods toolkit: summary

Figure 3, below, summarises the preceding discussion and sets out some of the main impact evaluation methods used in local economic development. The Figure uses the Scientific Maryland Scale (SMS) to rank the methodologies by level of robustness (or internal validity, to use more technical language). The SMS runs from Level 1 (minimum) to Level 5 (maximum). For reasons of space and focus I do not discuss these methods in detail; many technical guides are available, for example the UK Government Magenta Book (2020), which covers the full stack of impact and process evaluation tools, or the What Works Centre for Local Economic Growth (2016), which focuses on quantitative designs for impact evaluations. That said, four broader comments are worth making.

First, any given place-based or place-sensitive policy, our body of evidence will involve multiple methods, not just the most robust. This partly reflects the constraints on randomisation discussed above; it also reflects broader features of the policy and research landscape, which I discuss in Section 5. While we might prefer to use only RCTs and quasi-experiments (Heffetz and List 2021), for place-based policies this makes the best the enemy of the good. Practically we will need to work with many forms of evidence (Bates and Glennerster 2017).

Figure 3. Common methods for impact evaluation, ranked using the Scientific Maryland Scale.

Box 2: Our robustness scores (based on adjusted Maryland Scientific Methods Scale)

Level 1: Either (a) a cross-sectional comparison of treated groups with untreated groups, or (b) a before-and-after comparison of treated group, without an untreated comparison group. No use of control variables in statistical analysis to adjust for differences between treated and untreated groups or periods.

Level 2: Use of adequate control variables and either (a) a cross-sectional comparison of treated groups with untreated groups, or (b) a before-and-after comparison of treated group, without an untreated comparison group. In (a), control variables or matching techniques used to account for cross-sectional differences between treated and controls groups. In (b), control variables are used to account for before-and-after changes in macro level factors.

Level 3: Comparison of outcomes in treated group after an intervention, with outcomes in the treated group before the intervention, and a comparison group used to provide a counterfactual (e.g. difference in difference). Justification given to choice of comparator group that is argued to be similar to the treatment group. Evidence presented on comparability of treatment and control groups. Techniques such as regression and (propensity score) matching may be used to adjust for difference between treated and untreated groups, but there are likely to be important unobserved differences remaining.

Level 4: Quasi-randomness in treatment is exploited, so that it can be credibly held that treatment and control groups differ only in their exposure to the random allocation of treatment. This often entails the use of an instrument or discontinuity in treatment, the suitability of which should be adequately demonstrated and defended.

Level 5: Reserved for research designs that involve explicit randomisation into treatment and control groups, with Randomised Control Trials (RCTs) providing the definitive example. Extensive evidence provided on comparability of treatment and control groups, showing no significant differences in terms of levels or trends. Control variables may be used to adjust for treatment and control group differences, but this adjustment should not have a large impact on the main results. Attention paid to problems of selective attrition from randomly assigned groups, which is shown to be of negligible importance. There should be limited or, ideally, no occurrence of 'contamination' of the control group with the treatment.

Note: These levels are based on but not identical to the original Maryland SMS. The levels here are generally a little stricter than the original scale to help to clearly separate levels 3, 4 and 5 which form the basis for our evidence reviews.

Source: What Works Centre for Local Economic Growth (2016)

Second, though, while *all* of these designs can tell us something about programme impact, methods that score SMS3 and above give us much more confidence about policy effects than those below, because they are all closer to the ideal-type counterfactual design outlined in Section 3.1. By contrast, SMS2 designs use much less precise

methods to reconstruct the counterfactual, and SMS1 methods leave a lot of confounders in place. All of the examples in Box 1 are scored at either SMS3 or SMS4. This means that while we have to be flexible about what makes up our evidence base about a policy, we should still have a preference for more robust methods, both in existing bodies of evidence, and especially when commissioning new studies.

Third, while the Figure covers the most commonly used designs, many others are available (What Works Centre for Local Economic Growth 2016). For example, for large / one-off place-based interventions, no obvious counterfactual location may be available. In these cases, 'synthetic control' designs, can be used to build a control area from a weighted average of other locations (see Nathan (2022) for a recent example). Many other extensions to the basic difference in differences framework have also been developed in recent years (see Roth et al (2023) for an overview).

Fourth, when assessing existing evidence, SMS scores should ideally be applied to both evaluation design and its implementation. In the real world, evaluators often find that their plans need changing or refining on the fly, for example because of unforeseen problems in data collection; researchers may also overclaim for their papers, and/or make mistakes in execution.

3.3 / Data toolkit

Once we as evaluators have design ideas in place, we can turn to questions of data. Robust impact evaluation may be possible using existing / secondary datasets, but more often it will require a mix of primary, programme monitoring data alongside secondary data. Setting up effective monitoring systems and gathering rich information on programme inputs, outputs and outcomes is not only important in its own right, but is also essential for all kinds of evaluation, impact evaluation in particular. Relatedly, monitoring and evaluation design need to take place in parallel, in the early stages of project planning, so that we can be confident that monitoring systems are picking up the

information that will be needed for evaluation down the line. At the same time, policymakers need to make sure that this information is suitably cleaned and stored, to be easily available to evaluators later on. This happens much less frequently than one would think.

More broadly, recent years have seen two major developments in evaluators' data toolkit. First, in the UK and many other OECD member states, academic and other evaluators are increasingly able to access very large administrative microdata, giving rich information on people, households, firms and places over long time dimensions. This is an invaluable source of information for researchers in general, as well as for policy evaluation. Such highly structured 'administrative big data' is especially valuable when linked together, for example to form worker-firm panels or to combine information on individual/firm attributes with data on their surroundings. National governments and statistical agencies have important roles to play in continuing to gather such data and to make it available to research communities (see Nagaraj and Tranchero (2023) for evidence of the academic and policy impacts of secure data access in the US). Nevertheless, administrative data has some downsides: in particular, access is often complex and time-consuming for non-academics, so that policy analysts, especially at sub-national levels, will typically find it very challenging to use directly (see Section 4 for more on this). Furthermore, such data is inevitably backwards-looking, for example using industrial typologies that may be somewhat out of date, and relatively low-frequency, typically only available on an annual basis. While these features may not always matter for monitoring, or in an ex-post evaluation setting, they may sometimes place constraints on monitoring and evaluation efforts for place-based policies.

A second development, the explosive recent growth of online and other novel data sources, helps to deal with some of these constraints. Compared to conventional / administrative sources, 'Big Data' is typically higher frequency and offers higher dimensionality, often substantively so. Conversely, such data is often hard to access, can be expensive, often lacks detailed metadata, and – crucially – is usually implicitly sampled, creating bias if not corrected for (Einav and Levin 2014, Nathan and Rosso

2015). For these reasons, big data is best thought of as complementary to conventional data in the monitoring / evaluation toolkit; and policymakers should be alive to the additional cleaning / validation / quality control work that is typically needed to be able to use such data with confidence.

To my knowledge, these novel data are not yet widely used in policy evaluation within the public sector, although this will no doubt evolve in years to come. At the time of writing, academic big data use cases are mainly either in rich descriptive nowcasting / forecasting settings: for example using web-based indicators of firm activity to identify emerging industrial activities and clusters that aren't captured well in conventional data (Nathan and Rosso 2015, Kinne and Lenz 2019); using banking and other financial data to build local cost of living indices (Diamond and Moretti 2021); or using job ads to track and explain technology diffusion across places and firms (Calvino, Samek et al. 2022, Draca, Nathan et al. 2023). Notably, all of these studies combine administrative and novel data sources, using the first to complement and validate the second.

4 The state of the evidence

Parts 1 and 2 of this paper set out the case for robust evaluation of place-based policies, the key questions evaluators need to answer, and the problem of incomplete evidence. Part 3 gives an overview of the impact evaluation process. This part reviews the state of the evidence on our example place-based policies, and on local economic development programmes more broadly.

Figure 4. Summary of OECD-wide systematic reviews carried out by the What Works Centre for Local Economic Growth, 2013-2019.

Policy	# Studies	# SMS3+	Impact on jobs?	Positive
Access to Finance	1450	27	11	6
Apprenticeships	1250	27	9	7
Broadband	1000	16	10	5
Business Advice	700	23	17	8
Employment training	1000	71	65	33
Estate renewal	1050	21	5	1
Innovation	1700	63	10	6
Public realm	1140	0	0	0
Sports and culture	550	36	16	4
Transport	2300	29	6	2
Employment Zones	1300	30	27	15
EU Structural Funds (GDP)	1300	18	11*	5*

Figure 4, above, summarises the results of OECD-wide systematic reviews carried out between 2013 and 2019 by the What Works Centre for Local Economic Growth. The 12 reviews cover thousands of evaluations across a range of traditional place-based

policies, as well as place-sensitive policies targeted at people and firms. The first column shows the review topic, the second column the number of evaluations identified (of all types), the third column the number of impact evaluations meeting SMS3 or above on the Maryland Scale. The fourth and fifth columns show, respectively, the number of SMS3+ studies looking at employment effects, and the number finding positive impacts.

What can we learn from this figure? The first point is that, conditional on meeting the minimum standard, the evaluation evidence is arguably quite encouraging on the effectiveness of local economic development interventions in general, including place-based policies. For most policy areas, the Figure focuses on employment outcomes: here the median success rate is 50%. For EU Cohesion Policies, where evaluations focus on GDP/capita outcomes, the success rate is 45%.[11] If we look at other outcomes across reviews, we get similar results. Given the structural mega-forces forces driving spatial disparities, and the relatively small scale of most interventions, this seems – to this author at least – like a positive read-out. Box 2, below, gives the main thematic findings from evidence reviews in EU Cohesion Policies and broadband policy, our two example topics.

The second point is that while there is a great deal of evaluation evidence out there, as suggested in Part 2, much of it is incomplete or (for impact) non-robust. In particular, we can see that the share of evaluations meeting the What Works Centre's – demanding – minimum standard for counterfactual build is typically very small, compared to the larger body of evaluation evidence.

---

[11] The remaining studies largely consist of statistically insignificant effects of a policy (that is, where the estimated effect size cannot be distinguished from zero). Negative effects are extremely rare. Studies which find negative effects typically explore this and provide some high-level explanation: for example, poor implementation or extreme negative selection. In some cases negative effects need careful interpretation. For instance, accelerator programmes do not always have positive effects on participant firm survival, compared to observably similar non-participants. But this is usually the result of deliberate programme design: accelerators force founders to rigorously test the strength of their startup, setting up another venture if the underlying rationale is not strong enough.

Box 2: overall effectiveness of EU Cohesion Policies and broadband policies

EU Structural Fund support has a positive impact on regional GDP per capita in a little under half of the evaluations that consider GDP effects [11 studies]. Half of the studies which look at employment effects show a positive effect of EU support on employment [4 studies]. The evidence on a range of other outcomes is mixed (with only one study per outcome). Positive impact is bigger in relatively more developed regions, in four out of the five studies which consider this. Consistent with this, two out of three studies that consider the 'dose' (e.g. expenditure per capita) suggest an optimum 'level' of treatment [3 studies]. At the time of review, there was no evidence on the extent to which the different components of spend change the effectiveness of support.

Extending broadband internet to an area can affect firm productivity, number of businesses, and local labour market outcomes. These effects are not always positive, are not necessarily large, and may depend on complementary investments by firms (for example, training workers, or reorganising sales strategy or supply chains to take advantage of faster internet connections). Service industries and skilled workers appears to benefit more than manufacturing industries and unskilled workers. Economic effects tend to be larger in urban areas (or close to urban areas) than in rural areas. The effects of adoption and provision may differ, although far fewer studies look at the former than the latter. At the time of review, there were only three high quality evaluations of specific broadband policies (voucher schemes, direct public provision or public/private partnerships).

Drawn from What Works Centre for Local Economic Growth (2015, 2016).

Perhaps counter-intuitively, it turns out that this picture does *not* change very much if we relax the minimum standard. The What Works Centre ran internal exercises using an SMS2 standard, adding very few studies to sample reviews; more broadly, the Centre's

body of toolkits, which all use SMS2 minimum standard, typically found few studies at SMS2 level. Rather, the problem of missing or incomplete evaluation evidence is driven by three broader patterns. First, the What Works Centre most commonly found evaluations which focused on tracking and describing inputs and outputs, and explored process issues in detail, but did not discuss impact at all, or convincingly. Second, this was much more common than the converse: studies that robustly considered impact questions, but had little to say about important second order issues, especially on distributional effects and replication across places. Third, more common than both groups were ex-ante evaluations, which modelled the possible impacts of policies before they happened but did not follow up with explorations of actual implementation, or impact.

What explains these patterns? As we saw in Part 3, one set of issues stems from the fact that evaluation evidence, especially impact evaluation, is technically difficult to make and use: it is *hard to know*. A second set of issues arises from the institutional and cultural features of evidence makers and users, and what we might call the political economy of evidence; these dynamics mean evaluation evidence is often something *we don't want to know.*

5 Evaluation challenges: making and using evidence

Parts 1 and 2 of this paper set out the case for robust evaluation of place-based policies, the key questions evaluators need to answer, and the problem of incomplete evidence. Part 3 gives an overview of the impact evaluation process. Part 4 reviews the state of the evidence on our example place-based policies, and on local economic development programmes more broadly. This part considers a range of technical, institutional and cultural challenges that researchers and policymakers face in making and using evaluation evidence.

5.1 / Design challenges

One core difficulty in generating evaluation evidence is building robust research designs. Part 3 of the paper considers these issues in some detail for impact evaluation, perhaps the most challenging type of evaluation evidence to construct.

There are also other fundamental technical challenges at the evaluation design stage. One, perhaps surprisingly, is asking the right – or most useful – question. As Part 2 of the paper explains, one first order evaluation question to ask of a policy is 'does it work?', and this is the question that many (academic) evaluators start with. When monitoring and evaluating place-based policies, however, it is not always the most helpful starting point. As Mason et al (2023) argue, most policymakers do not start from the abstract position

of having a budget that can be spent or left; rather, they have resources that are pre-committed to a policy area, and need to know how best to spend it.[12]

A further technical issue concerns outcomes that are hard to observe because of timeframes. The long timeframe of many place-based policies puts constraints on what we can know. This is particularly the case for capital investments, like transport infrastructure, where the timeframe for wider economic impacts stretches over 50 years or more. However, it can also affect place-sensitive policies such as targeted active labour market programmes – where participants can get worse employment outcomes during the programme, and shortly afterwards before any policy effects kick in. As we shall see in the next section, the best response to these timing issues is a combination of well-grounded evaluation designs, and sequenced feedback systems that allow policymakers to track outputs, and see outcomes at multiple points in time, helping to build and maintain buy-in. This is especially important in very long-term Big Push-style policies, as in post-unification Germany (Enenkel 2021) or New Deal-era US rural development programmes (Kline and Moretti 2014).

A related challenge concerns outcomes that we care about but that are hard to measure quantitively and/or at scale. The obvious case here is wellbeing. As Sanders and Breckon (2023) point out, there are particular challenges in both how we can measure wellbeing for monitoring and evaluation, as well as questions about *when* we should be measuring it – retrospectively, or at the time a policy is running.

In the UK, there is now detailed guidance on wellbeing evaluation through the Government's Green Book (HM Treasury 2022) alongside metrics developed by the Office for National Statistics (2018). The 'ONS4' measures are summarised in Figure 5, below. Both institutions encourage practitioners to use revealed or stated preference methods to assess wellbeing outcomes of a policy: the former when we can use existing

---

[12] In addition, when reviewing evidence from a range of sources, a mixed evidence base – with some studies showing positive results, and others not – mean that simply asking 'what works?' is not always informative.

money-based proxies (such as population movements or property costs) and the latter when wellbeing questions need to be asked directly. There is now a growing body of evaluation evidence suggestion substantial wellbeing effects of local economic development programmes such as active labour market policies and green space provision (Franklin and Kenward 2023). Nevertheless, such tools require generating primary data, and are highly complex, requiring significant technical knowledge (Bakhshi, Fujiwara et al. 2015). I discuss capacity issues further below.

Figure 5. The UK's 'ONS4' wellbeing metrics.

Next I would like to ask you four questions about your feelings on aspects of your life. There are no right or wrong answers. For each of these questions I'd like you to give an answer on a scale of 0 to 10, where 0 is "not at all" and 10 is "completely".

| Measure	Question
Life Satisfaction	Overall, how satisfied are you with your life nowadays?
Worthwhile	Overall, to what extent do you feel that the things you do in your life are worthwhile?
Happiness	Overall, how happy did you feel yesterday?
Anxiety	On a scale where 0 is "not at all anxious" and 10 is "completely anxious", overall, how anxious did you feel yesterday?

Source: Office for National Statistics

Source: ONS (2018)

5.2 / Communication challenges

A second set of challenges focuses on messaging, and on packaging and communicating evidence from research to policymakers and practitioners. As Breckon and Sanders (2023) argue, evaluation evidence is often hard to use: too technically written, not explained clearly, and not made available in a timely enough fashion to get much policy traction.

At one level this is a written and visual communication problem, which can be tackled through non-technical language and infographics. Intermediary bodies like What Works Centres take on this synthesis and translation function, as do other bodies like thinktanks and JRCs. Figure 6, below, gives some examples of how a complex body of evaluation evidence can be distilled.

Figure 6. Example infographics summarising an evidence base.
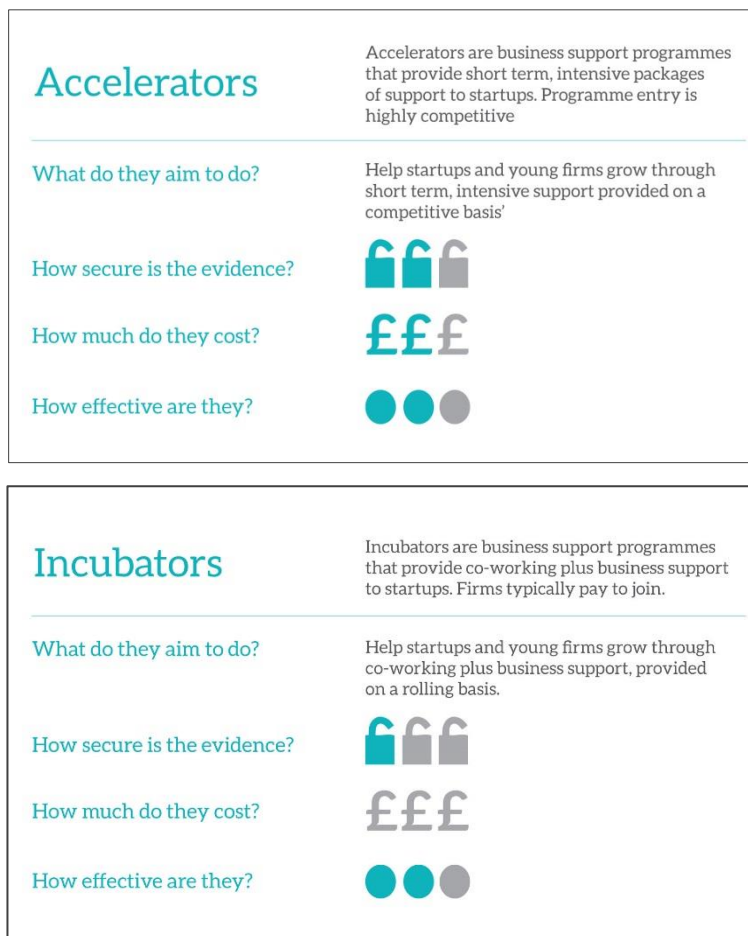
Source: What Works Centre for Local Economic Growth (2018)

One reason such translation work is necessary is that evidence generators and users typically exist in different institutional spaces, with distinct cultures and incentives, a point I return to below. More broadly, though, the diffuse nature of the place-based policy field presents particular challenges for researchers and policymakers looking to synthesise

evidence on place-based policies. In academic terms, 'place-based policy' is not a field or discipline, but rather the intersection of multiple fields and methodologies. In this it is much like geography, the body of knowledge that arguably sits at its core (Peet 1998). This structural constraint means that reviews of place-based programmes will almost never involve the formal meta-analysis practised in education, development or medicine. Here, a large body of studies using identical research designs is selected, and overall effects are derived from averaging impacts across individual studies. Rather, researchers looking at local growth need to use a form of realist synthesis (Pawson 2006) in which higher level patterns of results are combined with interpretation grounded in underlying theories of change. For example, researchers can develop typologies of impact (positive, zero, negative) and 'vote count' outcomes in each category (Card, Kluve et al. 2015). This strategy will only work, however, if the bodies of evidence in question are coherent, and can be scored for quality on criteria that make sense for that body of evidence.

5.3 / Capacity challenges

A third set of challenges relates to policymakers' and practitioners' capacity to use evidence generated. Many evaluation tools are technically complex, especially those relating to impact evaluation (see Part 3 of this paper); communicating these tools and their findings clearly is also not straightforward. As Mason et al (2023) point out, even *commissioning* a policy evaluation, and managing the evaluation effectively, is challenging: you need some expertise to be a good client. This includes not just the core business of ensuring a robust research design and its delivery, but also the principal-agent challenges of effective commissioning, contracting and oversight.

While member state national governments typically have some capacity to commission, use and understand these tools, this is not always a given, and even in quite devolved systems, the capacity problem is typically more acute at sub-national level. In countries like the UK, which have experienced extensive public sector cuts for the past decade, the capacity problem is severe at local level. Most concretely, as Mason et al point

suggest, evaluation presents a resource challenge: even commissioning and managing others' work has a financial and person-time cost, especially when things go wrong. Austerity also reduces public agencies' capacity to conduct monitoring and evaluation functions themselves; and perhaps just as significantly, reduces their capacity to act on evaluation evidence by reducing the set of feasible policy options.

5.4 / Institutional challenges

In theory, policy development is supposed to follow a strict 'policy cycle': develop a rationale, establish objectives, appraise options, monitor implementation, then evaluate and feedback (ROMAEF for short). Figure 7, below, gives an example of a policy cycle from the UK Government's Green Book.

However, in a series of conversations with officials and politicians, Hallsworth et al (2011) report that "virtually every interviewee dismissed policy cycles ... as being divorced from reality" (p5). Policymaking is often short term and event- driven; interventions are taken on before clear rationales are established; and the effects of policies are often indirect, diffuse, and take time to appear, making linear evaluation and feedback difficult. As one interviewee explains:

> *If you've got to be evidence-based, and inclusive, and joined up, and consultative, and outward-looking, you can't deliver a policy in a week – but ministers want policies tomorrow. (ibid p8)*

Figure 7. An example of a policy cycle.
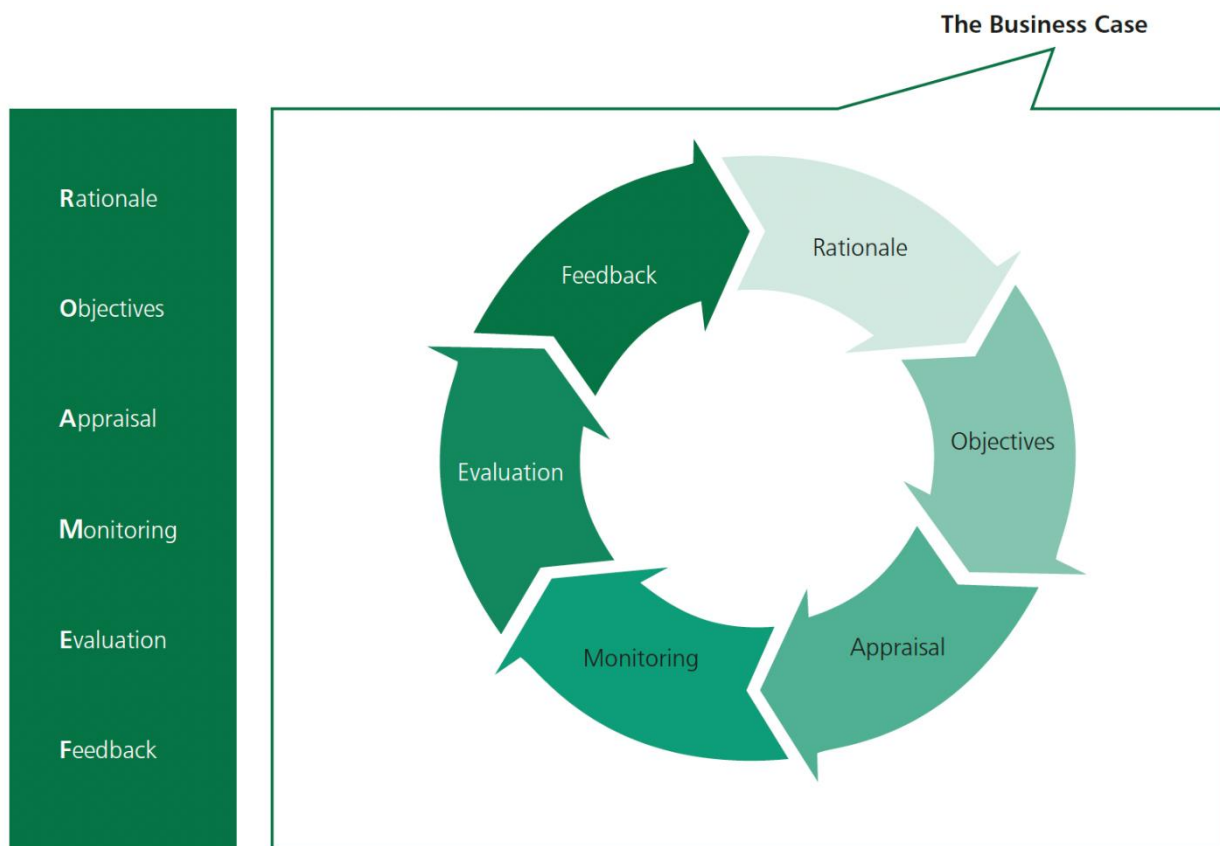
Source: HM Treasury (2022)

As we already know, in the real world, policymaking is more complex, and more nonlinear than we might like. All of this makes it challenging for policymakers and practitioners to make effective use of evidence, even when it is available and comprehensible, and where there is – in theory – capacity to use it. Lindblom (1959) makes a classic distinction between 'root' and 'branch' modes of policymaking. The first involves a comprehensive review of all available evidence, sifting for quality, and then taking a first-best decision that maximises some notion of welfare. The second involves a series of 'successive limited comparisons', drawing on what evidence is immediately available, and taking a second-best decision that minimises welfare tradeoffs. Lindblom suggests that that branch-based policymaking, or 'muddling through', is by far the most common.

Objectives are often unclear, conflicting and/or contested; underlying values and agendas conflict; actors have limited information and/or capacity to act on it.

Not only is the notion of the policy cycle out of sync with real-world practice, it arguably has a distorting effect on evidence-making and use. This is because – as Figure 7 shows – it presents evaluation as a self-contained element to be considered after implementation. One plausible effect of this, at least in the UK, is that policymakers put much more emphasis on making the ex-ante case for a policy, in order to get the resources to implement it, than on seeking to understand the (Bravo-Biosca 2019)ex-post effects of the policy after it has been rolled out.

As Mason et al (2023, p116) put it, 'policymakers are, perhaps understandably, more interested in figuring out ways to assess new projects, rather than discovering whether their estimates were right after the money was spent.' While this is short-term rational, policy cycle thinking creates an obvious long-term co-ordination problem. Rather, evidence and experience strongly suggest that not only does evaluation thinking have to built into the initial stages of policy design, but that policy implementation itself is often better served by a test / learn / adapt approach, combining exploration, pilots, robust evaluation and rapid scaling (Bravo-Biosca 2019). It is easy to see that this approach is fundamentally challenging for traditional place-based policy, for some of the same reasons that make RCTs hard to implement in this space (see Section 3 of this paper). Conversely, such an approach is suitable for deployment in many place-sensitive policy settings, where the immediate targets are people or firms in particular places, and where interventions are not particularly capital-intensive.

This discussion takes into deeper, more structural constraints: what we could call the political economy of evaluation, and specifically the different cultures and incentives of policymakers and (mainly academic) researchers, on top of the technical, communication and capacity issues discussed above. From the policymaker's perspective, there is considerable 'downside risk of being found to spend public money on something that

didn't work' (Mason et al 2023, p121). That is, policymakers rationally may not really want to know whether or not a policy is effective – particularly if they have already invested political capital in it, and/or are more concerned just to be seen to be doing something. Consistent with this, in experiments at World Bank and IADB workshops Vivalt and Colville (2023) find that compared to participants who are researchers, policymakers in the room are significantly more optimistic about likely effects of development policies. For similar reasons, officially commissioned evaluations of a policy will often look at a very wide range of outcomes. This may reflect a real attempt to do joined up policy, or reflect that a policy has a broad range of objectives. But it may also reflect what we can call "*O*-Hacking": a desire to mechanically increase the chances of finding a positive effect, even if this has little to do with the original objectives of the programme.[13]

From the researcher's perspective, the majority of academics select into a profession which is research orientated, and where career development generally depends on publications over engagement with policy, business and the wider public. Thus, writeups of academic evaluations often privilege the research design to the exclusion of any useful policy detail or policy implications, tendencies amplified by the space constraints of peer-reviewed journals. In turn, this makes these (often high quality) research findings of limited use to policymakers. While they do not present direct evidence, Sanders and Breckon (2023) raise the alternative scenario that researchers commissioned by government may find it hard to be fully independent in their evaluations: 'it is hard in practice to be openly critical of elected officials and the government of the day' [p138]. Both these forces limit evaluation evidence quantity and quality.

One very striking way in which these dynamics play out is in differences in the *type* of evaluation evidence that researchers and policymakers take more seriously. In further experiments, Vivalt et al (2023) found that academics place most emphasis on studies' internal validity – how robust the design is – policymakers strongly prefer evidence that generalises across locations, even if this is substantially lower quality. Crucially,

---

[13] Related to, but distinct from *P*-Hacking (see Head et al (2015)).

policymakers also placed a lot more weight on evidence recommended by known / local experts, even if this recommended less effective policies.

5.5 / Systemic challenges

These evaluation challenges come together, and are particularly pressing, when we think about systems transitions, the net zero transition in particular. As Section 1 set out, this urgent policy space has a significant place-based component, in particular through the major industrial policy programmes being rolled in out in the EU, US, China and elsewhere. However, compared to previous societal missions, the net zero mission is extremely broad, and the problem space is not fully defined (Robinson and Mazzucato 2019). In turn, this makes it hard for researchers and policymakers to develop theories of change around policies for the net zero transition, let alone designs for impact and process evaluation.

There are at least two complementary ways to tackle this challenge. At the micro level, there is a great deal of continuity in the policy instruments used to promote the net zero transition, and previous generations of industrial policy (Rodrik 2004, Tirole 2017). This continuity gives us both bodies of related evidence about how effective these programmes might be, both generically and – potentially – in closely related industrial / technology space. It also gives us a set of off-the-shelf research designs for impact and process evaluation, which can be adapted for specific net zero interventions.

At the macro level, powerful frameworks exist for thinking about socio-technical and socio-economic transitions and scenario-planning them at a broad level (Geels and Schot 2007, Perez 2010). The limitation on many of these frameworks is that they lack a clear spatial dimension, which makes it hard to plug in place-based policies. In response, some researchers have started thinking about missions have started applying those frameworks to specific local contexts (Flanagan, Uyarra et al. 2023). Overall, though, at macro level both the problem and solution spaces are less developed.

Conclusions / recommendations

Good monitoring and policy evaluation is extremely important to effective placed-based policymaking, but surprisingly hard to incorporate into mainstream policymaking infrastructures. This paper has set out the basic evaluation workflow, and the challenges to generating and using evaluation evidence. Monitoring and evaluation involves both practical challenges – things that are hard to know – but also cultural / institutional challenges, things we don't want to know.

How might we address these challenges, as researchers and policymakers? I conclude the paper by setting out some high-level suggestions.

First, framing. The previous section highlighted reasons why policymakers might want to avoid robust evaluation of the impacts their policies may (or may not) have had. These avoidance narratives are powerful and intuitive. However, one can also build an alternative narrative in which robust evaluation is a strength, rather than something to avoid. Policymakers adopting this stance would say that they are confident their policies are well-designed, drawing on existing evidence and best practice; are open to piloting and testing, to ensure that what has been effective elsewhere can also work in their particular context; and happy to lean from failure and to scale up success. A good recent example of this stance in practice is the approach that Greater Manchester, and some other UK city-regions, have adopted as part of their ongoing devolution settlements with UK central government. It should be noted that here, city leaders are deploying openness and rigour in policymaking and evaluation as a lever to extract greater powers and resources from national government. That is, they have a strong ulterior motive. Nevertheless, one can see how this positioning around evidence, and evaluation, might

also work in other contexts, especially ones in which city (or national) leaders are resource-constrained.

Second, rule-setting and incentives. Elected leaders, both at national and local level, can both act as 'champions' for evidence-based policy, as in the Greater Manchester example, *and* introduce policies that mandate the creation and use of robust evidence in policymaking. A crucial element in this policy mix is giving researchers access to data, especially microdata and information on programme creation / management (policy budgets, timelines, monitoring data, and for competitive funds, details of all applicants including losing bidders).

Similarly, grant-giving bodies, especially national governments and cross-national bodies like the EC, are able to set minimum standards for monitoring and evaluation, most obviously as a requirement for receiving funding. EU Cohesion Funds already do this; in the UK, national government is increasingly doing so, and using SMS-type tools to require robustness in impact evaluation, in particular. This rule-setting function can, in theory, change actors' incentives and tackle the co-ordination problems discussed above. However, rule-setting alone is not enough, unless communication and capacity problems are also addressed.

Third, information and communication. Monitoring and evaluation outputs are often extremely dry. Clearer language and use of visualisation is one simple way to make evaluation outputs more comprehensible and thus useful. Researchers also need to take seriously the idea that impact and process evaluations are complements, and that second-order questions for them are – from the policymaker's perspective – often the questions they care most about. Taking steps to make evaluations address process, context and distribution issues will help evidence become more useful and more likely to be taken on. In addition, appropriate messaging and framing of material can help; Vivalt and Colville (2023) find that compared to academics, policymakers are especially

appreciative of a) good news b) descriptive statistics and c) detail, all of which can trigger uptake of study results.

More broadly, how-to guides can help policymakers both implement effective evaluations, and (more usually) act as effective commissioners, and consumers of evaluation evidence. Guides developed by Bates and Glennerster (Bates and Glennerster) and the What Works Centre for Local Economic Growth (2022) are two good examples which combine ideas from experimentalist / RCT-based and realist / pluralist paradigms.   Bates and Glennerster (2017), for instance, set out principles for building an evidence base which emphasise the importance of theories of change, quality-weighting evidence, and using robust methods; but also the need to understand local context and conditions, combine different types of evidence, and adjust priors in the face of unexpected results.

Fourth, capacity and institutions. In the UK and some other member states, national and sub-national government capacity has been hollowed out in the past decade and a half. In a context where monitoring and evaluation is seen as secondary, these functions have often been the first to be depleted. Building these back up should be a priority: even if governments do not seek to do everything in-house, being an effective client to external consultants requires some experience and expertise. In the UK, the experience of the What Works Network is that this kind of institutional re-orientation is a 10-year plus process, which implies committing to long term objectives and resourcing these accordingly. Another crucial element in state capacity is systems to safely store, access and work with administrative microdata, including data linkage. Many OECD member states have some form of secure infrastructure for such granular data, but there is huge variation in ease / cost of access.

Finally, who should do this? Sanders and Breckon (2023), reflecting on the UK's experience, highlight the 'importance of co-ordinated, whole-systems thinking on cross-cutting policy matters' [p130]. External bodies like the OECD, and intermediaries like

What Works Centres can help here through convening actors, synthesising and communicating evidence, and influencing. But they rarely have statutory powers, and ultimately need to be seen as complements to the main players: national governments and international bodies like the EC play central roles in these kinds of actions, given their rule-setting, deep resources and ability to leverage economies of scale and scope.

References

Ahlfeldt, G., P. Koutroumpis and T. Valletti (2017). "Speed 2.0: Evaluating Access to Universal Digital Highways." Journal of the European Economic Association **15**(3): 586-625.

Akerman, A., I. Gaarder and M. Mogstad (2015). "The Skill Complementarity of Broadband Internet " Quarterly Journal of Economics **130**(4): 1781–1824.

Autor, D. (2019). "Work of the Past, Work of the Future." American Economic Review Papers and Proceedings **109**: 1-32.

Bailey, D., C. N. Pitelis and P. R. Tomlinson (2023). "Place-based industrial and regional strategy – levelling the playing field." Regional Studies **57**(6): 977-983.

Bakhshi, H., J. Edwards, S. Roper, J. Scully, D. Shaw, L. Morley and N. Rathbone (2013). Creative Credits: A randomised controlled industrial policy experiment. London, NESTA.

Bakhshi, H., D. Fujiwara, R. Lawton, S. Mourato and P. Dolan (2015). Measuring Economic Value in Cultural Institutions. London, AHRC.

Bates, M. A. and R. Glennerster (2017). "The Generalizability Puzzle." Stanford Social Innovation Review **15**(50-54).

Becker, S. O., P. H. Egger and M. von Ehrlich (2010). "Going NUTS: The effect of EU Structural Funds on regional performance." Journal of Public Economics **94**(9): 578-590.

Becker, S. O., P. H. Egger and M. von Ehrlich (2012). "Too much of a good thing? On the growth effects of the EU's regional policy." European Economic Review **56**(4): 648-668.

Becker, S. O., P. H. Egger and M. von Ehrlich (2013). "Absorptive Capacity and the Growth and Investment Effects of Regional Transfers: A Regression Discontinuity Design with Heterogeneous Treatment Effects." American Economic Journal: Economic Policy **5**(4): 29-77.

Bosquet, C. and H. G. Overman (2019). "Why does birthplace matter so much?" Journal of Urban Economics **110**: 26-34.

Brandily, P., M. Distefano, H. Donnat, I. Feld, H. G. Overman and K. Shah (2022). Bridging the gap: What would it take to narrow the UK's productivity disparities? London, Resolution Foundation.

Bravo-Biosca, A. (2019). Experimental Innovation Policy. NBER Working Paper 26273. Cambridge, MA, NBER.

Calvino, F., L. Samek, M. Squicciarini and C. Morris (2022). Identifying and characterising AI adopters. Paris, OECD.

Card, D., J. Kluve and A. Weber (2015). What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations. NBER Working Paper 21431. Cambridge, MA, NBER.

Cartwright, N. (2009). "How To Do Things With Causes." Proceedings and Addresses of the American Philosophical Association **83**(2): 5-22.

Cheshire, P., M. Nathan and H. Overman (2014). Urban Economics and Urban Policy. Cheltenham, Edward Elgar.

Coyle, D. (2021). Cogs and Monsters: What Economics Is, and What It Should Be. Princeton, Princeton University Press.

Crawford, N. (2023). "Europe's Measured Response to the US Inflation Reduction Act." Survival Online https://www.iiss.org/online-analysis/survival-online/2023/04/europes-measured-response-to-the-us-inflation-reduction-act/.

Criscuolo, C., R. Martin, H. G. Overman and J. Van Reenen (2019). "Some Causal Effects of an Industrial Policy." American Economic Review **109**(1): 48-85.

Diamond, R. and E. Moretti (2021). Where is the Standard of Living the Highest? Local Prices and the Geography of Consumption. Mimeo.

Draca, M., M. Nathan, V. Nguyen, J. Oliviera-Cunha, A. Rosso and A. Valero (2023). The New Wave: Technology diffusion in the UK during the 2010s, Mimeo.

Ehrlich, M. v. and H. G. Overman (2020). "Place-Based Policies and Spatial Disparities across European Cities." Journal of Economic Perspectives **34**(3): 128-149.

Einav, L. and J. D. Levin (2014). "The Data Revolution and Economic Analysis." Innovation Policy and the Economy **14**(1-24).

Enenkel, K. (2021). "What can German reunification teach the UK about levelling up?" Centre for Cities Blog https://www.centreforcities.org/blog/what-can-german-reunification-teach-the-uk-about-levelling-up/.

Fajgelbaum, P. D. and C. Gaubert (2020). "Optimal Spatial Policies, Geography, and Sorting*." The Quarterly Journal of Economics **135**(2): 959-1036.

Flanagan, K., E. Uyarra and I. Wanzenböck (2023). "Towards a problem-oriented regional industrial policy: possibilities for public intervention in framing, valuation and market formation." <u>Regional Studies</u> **57**(6): 998-1010.

Franklin, J. and T. Kenward (2023). Guide to assessing wellbeing impacts in the charity sector. London, Pro Bono Economics / What Works Centre for Wellbeing.

Geels, F. W. and J. Schot (2007). "Typology of sociotechnical transition pathways." <u>Research Policy</u> **36**(3): 399-417.

Glaeser, E. and J. D. Gottlieb (2008). "The Economics of Place-Making Policies." <u>Brookings Papers on Economic Activity</u> **2008**(1): 155-239.

Green, J. and R. De Geus (2022). Red Wall, Red Herring? Economic Insecurity and Voting Intention in Britain. Oxford, Nuffield Politics Research Centre.

Hallsworth, M., S. Parker and J. Rutter (2011). Policymaking in the real world: evidence and analysis London, Institute for Government.

Head, M. L., L. Holman, R. Lanfear, A. T. Kahn and M. D. Jennions (2015). "The Extent and Consequences of P-Hacking in Science." <u>PLOS Biology</u> **13**(3): e1002106.

Heffetz, O. and J. List (2021). "Who's Afraid of Evidence-Based Policymaking." <u>Artefactual Field Experiments</u> **00739**(The Field Experiments Website).

HM Treasury (2020). Magenta Book Annex A: Analytical methods for use within an evaluation. London, UK Government.

HM Treasury (2020). Magenta Book: Central Government guidance on evaluation. London, UK Government.

HM Treasury (2022). The Green Book: Central Government guidance on Appraisal and Evaluation. London, HM Treasury.

Kinne, J. and D. Lenz (2019). Predicting Innovative Firms Using Web Mining and Deep Learning. <u>ZEW Discussion Paper No. 19-001</u>. Mannheim, ZEW.

Kline, P. and E. Moretti (2013). People, Places and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs. <u>NBER Working Paper 19659</u>. Cambridge, MA, NBER.

Kline, P. and E. Moretti (2014). "Local Economic Development, Agglomeration Economies and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority." <u>The Quarterly Journal of Economics</u> **129** (1): 275-331.

Lindblom, C. (1959). "The Science of 'Muddling Through'." <u>Public Administration Review</u> **19**(2): 79-88.

Mason, D., M. Nathan and H. Overman (2023). The What Works Centre for Local Economic Growth: some lessons from the first ten years. The What Works Centres: Lessons and Insights from an evidence movement. M. Sanders and J. Breckon. Bristol, Policy Press.

Mazzucato, M. (2018). "Mission-oriented innovation policies: challenges and opportunities." Industrial and Corporate Change 27(5): 803-815.

McCann, P. (2023). How Have Place-Based Policies Evolved to Date and What Are They For Now? Paris, OECD.

McCann, P., R. Ortega-Argilés, D. Sevinc and M. Cepeda-Zorrilla (2023). "Rebalancing UK regional and industrial policy post-Brexit and post-Covid-19: lessons learned and priorities for the future." Regional Studies 57(6): 1113-1125.

Molloy, R., C. L. Smith and A. Wozniak (2017). "Job Changing and the Decline in Long-Distance Migration in the United States." Demography 54(2): 631-653.

Moretti, E. (2022). Placed-Based Policies and Geographical Inequalities. IFS Deaton Review of Inequalities Working Paper. London, IFS.

Muro, M., J. Parilla and M. Ross (2023). What state and local leaders need to know about the federal government's regional Tech Hubs competition. Washington DC, Brookings Institution.

Nagaraj, A. and M. Tranchero (2023). How Does Data Access Shape Science? Evidence from the Impact of U.S. Census's Research Data Centers on Economics Research. NBER Working Paper 31372. Cambridge, MA, NBER. No. 31372.

Nathan, M. (2015). Microsolutions for Megaproblems: What works in urban regeneration policy? . After Urban Regeneration: People, communities and place. P. Matthews and D. O'Brien. Bristol, Policy Press.

Nathan, M. (2016). Linking Research and Policy in Local Economies. A Research Agenda for Regeneration Economies. L. Andres, J. Bryson and R. Mulhall. Cheltenham, Edward Elgar.

Nathan, M. (2022). "Does light touch cluster policy work? Evaluating the tech city programme." Research Policy 51(9): 104138.

Nathan, M. (2022). "The Long and Winding Road." Medium https://maxnathan.medium.com/the-long-and-winding-road-1f6a77028d05.

Nathan, M. and A. Rosso (2015). "Mapping digital businesses with Big Data: some early findings from the UK " Research Policy 44(9): 1714-1733.

Neumark, D. and H. Simpson (2015). Place-Based Policies. Handbook of Regional and Urban Economics. G. Duranton, J. V. Henderson and W. C. Strange, Elsevier. 5: 1197-1287.

OECD (2022). Addressing territorial disparities in future infrastructure needs in the wake of the COVID-19 crisis Paris, OECD.

Office for National Statistics (2018). Personal Wellbeing User Guidance. Newport, ONS.

Overman, H. and X. Xu (2022). Spatial Disparities Across Labour Markets. IFS Deaton Review of Inequalities Working Paper. London, IFS.

Pawson, R. (2006). Evidence Based Policy: A realist perspective London, Sage.

Peet, R. (1998). Modern Geographical Thought. Oxford, Blackwell.

Pellegrini, G., F. Terribile, O. Tarola, T. Muccigrosso and F. Busillo (2013). "Measuring the effects of European Regional Policy on economic growth: A regression discontinuity approach." Papers in Regional Science **92**(1): 217-233.

Perez, C. (2010). "Technological revolutions and techno-economic paradigms." Cambridge Journal of Economics **34**: 185-202.

Reeves, R. (2023). "Securinomics". London, Labour Party.

Robinson, D. K. R. and M. Mazzucato (2019). "The evolution of mission-oriented policies: Exploring changing market creating policies in the US and European space sector." Research Policy **48**(4): 936-948.

Rodríguez-Pose, A. (2018). "The revenge of the places that don't matter (and what to do about it)." Cambridge Journal of Regions, Economy and Society **11**(1): 189-209.

Rodrik, D. (2004). Industrial Policy for the Twenty-First Century. CEPR Discussion Paper 4767. London, Centre for Economic Policy Research.

Roth, C., P. H. C. Sant'Anna, A. Bilinski and J. Poe (2023). What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature. https://arxiv.org/abs/2201.01194v3.

Sanchis-Guarner, R., J. Montalbán and F. Weinhardt, (2021) (2021). Home Broadband and Human Capital Formation CESifo Working Paper No. 8846. Munich, CESifo.

Sanders, M. and J. Breckon (2023). Criticisms and Challenges of the What Works Centres. The What Works Centres: Lessons and insights from an evidence movement. M. Sanders and J. Breckon. Bristol Policy Press.

Sanders, M. and J. Breckon (2023). The Scene is Set. The What Works Centres: Lessons and Insights from an evidence movement. M. Sanders and J. Breckon. Bristol, Policy Press**:** 3-11.

Suedekum, J. (2021). Place-Based Policies - How to Do Them and Why. CEPR Discussion Paper No. DP16468. London, CEPR.

Suedekum, J. (2023). How Have Place-Based Policies evolved to date and what are they for now? . Paper prepared for EC-OECD seminar series 'Place-based policies for the future'. Paris, OECD.

Sullivan, J. (2023). Remarks on Renewing American Economic Leadership. Washington DC, US Government.

Tirole, J. (2017). Economics for the Common Good. Princeton, Princeton University Press.

Tooze, A. (2022). Welcome to the world of the polycrisis. Financial Times. 28 October.

Tooze, A. (2023). "Chartbook 220: Biden's "new industrial policy": Revolution in the making, or an exercise in defying gravity? ." Chartbook https://adamtooze.substack.com/p/chartbook-220-bidens-new-industrial.

Tunstall, R. and R. Lupton (2003). Is Targeting Deprived Areas an Effective Means to Reach People? An Assessment of One Rationale for Area-Based Funding Programmes. CASEpaper 70. London, London School of Economics.

Vivalt, E. and A. Colville (2023). How Do Policymakers Update Their Beliefs? Mimeo.

Vivalt, E., A. Colville and K. Sampada (2023). Weighing the Evidence: Which Studies Count? Mimeo.

What Works Centre for Local Economic Growth (2015). Evidence Review 6: Broadband. London. WWCLEG.

What Works Centre for Local Economic Growth (2016). Evidence Review 10: Area-Based Initiatives: EU programmes. London, WWCLEG.

What Works Centre for Local Economic Growth (2016). Guide to Scoring the Evidence using the Scientific Maryland Scale. London, WWCLEG.

What Works Centre for Local Economic Growth. (2022). "An 8-step guide to better evaluation." Retrieved 10 August, 2023, from https://whatworksgrowth.org/resource-library/8-step-evaluation-guide/.