what works centre for local economic growth

Undertaking impact evaluation of devolution

Introduction

Central and local government want to understand the impact of their policies. The most robust way of establishing this is through impact evaluation, as this considers what would have happened if the policy had not been pursued (known as the counterfactual).

Both the Ministry for Housing, Communities, and Local Government (MHCLG) and individual mayoral combined authorities (MCAs) have expressed an interest in evaluating the overall impact of devolution. An impact evaluation would assess whether devolution had an impact on (that is, caused a change in) outcomes for an area that had received devolved powers and funding compared to similar areas that had not. Alternatively, an evaluation could look at the impact of devolution across multiple MCAs.

This briefing aims to help policymakers think through the feasibility of evaluating the impact of devolution.

Background

Since 2014, 11 English city-regions have agreed one devolution deal or more.[1] Devolution deals involve establishing and transferring a range of powers, programmes, and budgets to MCAs. The powers vary across MCAs (with longer established MCAs generally having more powers) and can include adult education, skills, transport, business support, and housing. MCAs also receive a funding settlement. Initial devolution deals are agreed through a negotiation process between the UK government and local authorities, with subsequent rounds of negotiation also involving the (now established) MCA.

The Labour Government elected in July 2024 is committed to devolution, with an English Devolution Bill included in the King's Speech and the government writing to local authorities in the early weeks of

1 This does not include the Greater London Authority as it has a different legal basis – but the issues discussed would also apply to it.

the new parliament to encourage them to pursue devolution deals.[2, 3]

Evaluating devolution

Establishing a counterfactual

Impact evaluation helps us understand whether outcomes can be attributed to an intervention. For example, did training help an unemployed person gain employment. This is known as 'causal impact' or 'causality'– because it establishes that the intervention caused the outcome. Impact evaluations establish causality by using comparison. The ideal scenario would be to compare the same individual, business or place under two different scenarios – one where they receive support (known as 'treatment') and one where they do not. As establishing this counterfactual is not possible, an alternative comparison is needed. This is done by establishing a control group or comparison group. The standard approach is to create a group of individuals, businesses or places that are similar to those being treated, but that did not receive treatment. Changes in outcomes can then be compared between the treatment group and the comparison group.

Comparing groups that are similar helps address the possibility that factors other than the interventions might be causing changes. For example, when considering the impact of a training course, one concern may be that those that take up training are more motivated than those that do not and that this (rather than the training) may be responsible for them moving into work. The main challenge in impact evaluation is how to ensure the comparison group is similar to the treatment group. Various approaches are available and the feasibility of using each to evaluate devolution is considered below.[4]

Given the focus on evaluating the impact of devolution as a whole, the unit of observation will be the 'area', with the treated group (MCA or MCAs) compared to areas that are similar but do not have a devolution deal.[5] For this reason, when considering each of the approaches below, the focus is on whether it is possible to establish a comparison group of similar areas.

Randomisation

Randomised control trials involve randomly allocating eligible participants to either the treatment or comparison groups. This approach is considered the 'gold standard' of impact evaluation. In relation to devolution, this would mean some eligible areas randomly being allocated devolved powers and funding and others not. However, devolution deals have not been, and will not be, allocated on a random basis meaning this approach cannot be used to evaluate devolution as a whole.

Some evaluations exploit randomness that was not deliberately imposed but occurred for some other reason.

- The ***instrumental variables*** method involves finding something that explains treatment but has no direct effect on the outcome of interest (and is not related to any other factors that might determine that outcome). There are no obvious 'instruments' that could enable the use of this method to evaluate devolution as a whole.[6]

2 https://www.gov.uk/government/speeches/the-kings-speech-2024
3 https://www.gov.uk/government/publications/letter-from-the-deputy-prime-minister-to-local-leaders-the-next-steps-to-devolution/letter-from-the-deputy-prime-minister-to-local-leaders-the-next-steps-to-devolution
4 More information on these methods is available in our guide to scoring the evidence.
5 This also applies if the unit of analysis is individuals or businesses. In these cases, individuals or businesses within the MCA and comparison areas would be compared.
6 Whilst there are no instruments obvious to What Works Growth or those we have consulted with, there may be an instrument and dedicating some limited resources to exploring this option may be valuable.

- The ***regression discontinuity design*** method can be used in cases where there are cut-offs for treatment eligibility (for example, if areas with GVA per head less than English average were given devolution). This approach exploits the fact that whilst treated units are likely to be different to untreated units, the units close to the cut-off are likely to be similar in terms of observable and unobservable characteristics. As devolution is not determined based on cut-offs against standard criteria, this approach cannot be used to evaluate devolution as a whole.

Comparing similar areas

An alternative approach is to compare the treated group against a comparison group based on observable characteristics, before-and-after the treatment date.

- The ***difference-in-differences*** method evaluates the impact of an intervention by first calculating the change in the outcome for the treated group, and then subtracts the change in the outcome for the comparison group over the same period. This approach is better than a simple before and after comparison for treated areas, because it accounts for the fact that changes in outcome can be due to many different factors and not just the treatment. In addition, because it subtracts the differences both before and after the treatment, it controls for any unobserved differences between the two groups that do not change over time. However, it does not account for unobservable differences that vary with time. Whilst in most cases, the comparison areas would be 'real' geographies, one option is to establish a ***synthetic control*** area by weighting outcomes for a number of different geographies. Weights are chosen to ensure the synthetic control is as similar as possible to the treated area in the run up to treatment (in this case, in the run up to devolution).[7]
- ***Panel data methods*** use data that follow the same units over time allowing the evaluator to control for things that remain constant for each unit across time. Within this, the ***fixed effects approach*** uses dummy variables for individual characteristics and anything that happened in a given time period, whilst the ***first differences approach*** subtracts the outcome of a previous period in the final regression cancelling out anything that remains constant across time.[8] Panel data methods successfully control for observable and unobservable characteristics that remain constant throughout time but do not account for unobservable characteristics that change with time. Difference-in-difference can sometimes be implemented using panel data methods.[9]

For both approaches, the main challenge is finding a suitable comparison area from among non-devolved areas. The first two areas to agree devolution deals (and those with the most advanced devolution deals) were Greater Manchester and West Midlands – the largest city regions outside London. Most other devolved areas are city regions, with most major cities in the North of England and many in the Midlands now covered by a devolution deal. This means there are no obvious areas that could act as comparators.

7 A range of extensions to these methods – including staggered difference-in-differences and augmented synthetic control method – are also available but, as these are still being developed (for example, the academic literature has not yet settled on which of one of half a dozen or more staggered difference-in-difference estimators is most efficient), we do not recommend their use by central or local government at this stage to answer this question. However, these may become viable options for evaluating devolution in the future.

8 When using the fixed effects method, the fixed effects must be at the unit of analysis (i.e. if the unit of analysis is the area, the fixed effects must refer to the area, whereas if the unit of analysis is businesses, the fixed effects must refer to businesses).

9 Another possible option is to use Heckman two-stage correction (H2S) with difference-in-differences or panel data. Heckman two-stage corrections can also be used with instrumental variables and cross-sectional data (discussed elsewhere in this paper). However, as H2S have to meet the criteria for these methods alongside those specific to H2S, the same issues are likely to apply.

One potential option would be to exploit differences in the timing, using areas that are treated later (i.e. agreed a devolution deal later) as a comparison for areas that are treated earlier. Although devolution deals have been agreed at different times, the newly devolved areas are different to those agreed earlier. The differences will be even greater for future devolution deals as these areas will generally be smaller and less urban. Again, this makes it difficult to implement this approach. Another timing issue if difference-in-differences approach is being used is that changes in outcomes in areas treated 'early' must happen before those treated later receive treatment (in this case, devolution). Many of the outcomes devolution is aiming to affect will take a long time to change, limiting the potential use of this approach, as will the possibility of an MCA area agreeing multiple devolution deals with government over time.

One final option is to compare areas with different intensities of treatment. For example, comparing devolved areas receiving different amounts of funding (or somehow quantifying the 'strength' of powers devolved) to those receiving less. This approach is only appropriate if the intensity varies across areas in a way that is independent of the outcomes of interest. Although funding and powers vary across the MCAs, this is a result of negotiation, and is influenced by a range of factors including population size, economic performance, track record of delivery, and need. As these factors may also relate to the outcomes of interest, comparison across places using variation in intensity would not be appropriate.

An additional challenge in using difference-in-differences is that it requires the time period for the intervention to be known. Whilst the dates of devolution deals are known, timescales for when funding and powers were used are less clear-cut.

In summary, the process for negotiating devolution deals makes it difficult to identify similar areas that can be used as a comparison group. This makes it unlikely these approaches can be used to evaluate the overall impact of devolution. There is potential to use them in the future if devolution policy was designed to reflect them (for example, if future devolution deals were prioritised using cut-offs or if there was phased roll-out of new deals). These approaches could also potentially be used to evaluate specific elements of existing devolution deals.

Other methods

All methods that score SMS 3 or above on the Maryland Scientific Methods Scale have been considered above. Given that none are likely to be feasible for evaluating devolution as a whole, MHCLG and MCAs may wish to consider less robust methods that score SMS 2.

- ***Cross-sectional regressions*** compare treated units with untreated units. This requires a dataset that features many different units at one point in time. As such, it is not likely to be feasible for comparing areas (as there are too few) but it may be feasible if the unit of analysis was businesses or individuals. For this method to score SMS 2 it requires the use of control variables to take account of the different characteristics of the treated and comparison groups that may influence the outcome. The feasibility of this approach is therefore dependant on the ability to identify and control for these observable characteristics.
- ***Before-and-after method*** uses a time series dataset for which one unit is tracked across time, including both before and after treatment. The challenge with this approach is that other contextual factors and the unit's characteristics may affect the outcome in addition to treatment. Again, relevant control variables must be used for this approach to score SMS 2, meaning its feasibility will depend on the ability to identify and control for these.

Both approaches are potentially feasible and MHCLG or MCAs may wish to pursue them. However, as these are less robust methods, we encourage careful consideration of whether they would provide

convincing answers to key policy questions, and if they are pursued, to minimise the resources (time and money) dedicated to them. Pursuing other options – such as improved monitoring or evaluating specific elements of devolution deals – may be a better use of resources.

Theory-based evaluation

The UK government's guide to evaluation, the Magenta Book, also considers theory-based impact evaluations. This approach draws conclusions about an intervention's impact by assessing "*whether the causal chains thought to bring about change are supported by sufficiently strong evidence and that alternative explanations can be ruled out*".[10] Theory-based evaluations require a well-developed theory of change to be in place.

Theory-based evaluations are often proposed when a counterfactual impact evaluation is not possible. It has been suggested that they may be useful in complex policy environments or systems, when a combination of interventions are being pursued or the interventions may change over time, and in scenarios where the transferability of results to other contexts is needed. All of these would apply to devolution.

The rigour of theory-based evaluations relies on:

- Coherence of the theories
- Evidence that is specific enough to test the theories
- Triangulation of multiple sources
- Ruling out of alternative causes in order to claim impact
- Critical reflection and opening up to peer review and external scrutiny.[11]

Given that there are so many potential pathways by which devolution can affect high-level outcomes of interest (such as employment, health or poverty), it is unclear how these can be rigorously tested using a theory-based approach. As with other methods, it should be possible to undertake theory-based evaluation of specific interventions, although when counterfactual impact evaluation methods are feasible, these should be used rather than theory-based methods if the interest is in establishing the causal impact of policy.

Other challenges

Whilst the lack of a suitable comparison areas is a key challenge in implementing a robust counterfactual impact evaluation of devolution as a whole, there are also several other challenges that would make it difficult even if a counterfactual could be identified.

- Given the breadth of policy areas covered by devolution, each MCA will be attempting to affect multiple economic, social, and environmental outcomes. The breadth of policy areas can make identifying the appropriate outcomes difficult and can make it more difficult to construct treatment groups where nothing is happening on one or more of these areas. This also makes data collection and analysis for evaluation more complex and time-consuming and raises problems with multiple hypothesis testing (which tends to find an effect on one of many outcomes, even when the policy has no impact on any outcome). These issues would be compounded in an evaluation that looked at the impact of multiple MCAs as different MCAs may be using the same devolved powers to achieve different outcomes and the exact mix of policies and outcomes will vary. This makes it hard to look at outcomes over multiple areas, limiting the extent to which findings are generalisable.

10 HM Treasury (2020). Magenta Book. Central government guidance on evaluation.
11 HM Treasury (2020). Magenta Book. Central government guidance on evaluation.

- The importance of the policy in affecting outcomes, relative to everything else that affects the outcome, will also matter. Where a policy only plays a small role, it is less likely that an effect will be detected. As many of the outcomes that devolution seeks to address – such as employment, health, poverty, and carbon emissions – are complex, a range of other factors and wider events that will affect these, making it more difficult to establish the impact of devolution.
- As impact evaluation relies on statistical methods, the number of observations matters. More observations is better, increasing the statistical 'power' – the likelihood of getting statistically significant results which accurately reflect whether the policy has positive, negative or no effects. This is less likely to be an issue for evaluations where the unit of analysis is individuals or businesses as there will be large numbers in the treated and comparison groups. However, if evaluation uses the MCA as the unit of analysis, there will be few observations, making it less likely that the evaluation will detect effects. This is a challenge as many of the outcomes that policymakers are likely to be most interested in are at measured at the area-level.
- MCA geographies are decided on an *ad hoc* basis by local policymakers when pursuing devolution deals (with UK government sometimes requesting variations during the negotiation). MCAs always cover multiple local authority areas, and generally do not reflect previously defined administrative or statistical geographies. If an impact evaluation was pursued, this would mean comparison groups would need to be defined, with the basis for these geographies subject to debate.

Overall assessment and alternatives

Overall, our assessment is that it is likely to be extremely difficult to evaluate devolution as a whole due to how the policy has been designed and implemented over time. Whilst there may be some options available, they are likely to either use less robust methods or not provide convincing answers to key policy questions.

Even though evaluation of devolution as a whole is unlikely to be possible, it should be possible to undertake impact evaluation of some policies, programmes or projects delivered by MCAs as a result of funds or powers received through a devolution deal. This could include:

- Evaluating single interventions within a programme.
- Evaluating multiple interventions with a shared outcome.
- Evaluating multiple interventions that targeted the same geographic area within a devolved area.

There is scope to use a wide range of counterfactual impact evaluation methods, including randomised control trials, that could be used to evaluate specific interventions, especially if evaluation is built in from the design phase.

In addition, MHCLG and MCAs could learn more about how devolution is influencing practice by analysing monitoring data and undertaking process evaluation. For example, monitoring data and process evaluation could help MHCLG understand how funding and powers have been used by each MCA and how this is different to pre-devolution period. This would be helpful as this would form the first step in any theory of change about how devolution impacts on outcomes.

what works centre for
local economic growth

This work is published by the What Works Centre for Local Economic Growth, which is funded by a grant from the Economic and Social Research Council, the Department for Business and Trade, the Ministry of Housing, Communities and Local Government, and the Department for Transport. The support of the Funders is acknowledged. The views expressed are those of What Works Growth and do not represent the views of the Funders.

November 2024

What Works Centre for Local Economic Growth

info@whatworksgrowth.org
@whatworksgrowth

www.whatworksgrowth.org

HM Government

UKRI Economic and Social Research Council