



INNOVATION &
RESEARCH
CAUCUS

AN INTEGRATED APPROACH TO ESTIMATING THE IMPACT OF UKRI

IRC Insight Paper 012

REPORT PREPARED BY

Adam Brown

Cambridge Econometrics



Delivered with
ESRC and
Innovate UK

Contents

An Integrated Approach to Estimating the Impact of UKRI.....	3
Theory of Change and Systems Mapping	3
Map Development Process	5
Systems Map content	6
Hypotheses Development and Econometric Approaches.....	8
Combining Systems Mapping and Econometric Approaches	9
References	13
Appendix.....	14
Author-specific questions:.....	15

Authors

The core members of the research team for this project were as follows:

» Dr Adam Brown - Cambridge Econometrics

This document relates to IRC Project FFEI0004 – Evidencing Impact Projects

Acknowledgements

This work was supported by Economic and Social Research Council (ESRC) grant ES/X010759/1 to the Innovation and Research Caucus (IRC) and was commissioned by UK Research & Innovation (UKRI).

We are very grateful to the project sponsors at UK Research and Innovation (UKRI) for their input into this research. The interpretations and opinions within this report are those of the authors and may not reflect the policy positions of UKRI.

We would also like to acknowledge and appreciate the efforts of the IRC Project Administration Team for their support in preparing this report.

About the Innovation and Research Caucus

The Innovation and Research Caucus supports the use of robust evidence and insights in UKRI's strategies and investments, as well as undertaking a co-produced programme of research. Our members are leading academics from across the social sciences, other disciplines and sectors, who are engaged in different aspects of innovation and research systems. We connect academic experts, UKRI, IUK and the ESRC, by providing research insights to inform policy and practice. Professor Tim Vorley and Professor Stephen Roper are Co-Directors. The IRC is funded by UKRI via the ESRC and Innovate UK, grant number ES/X010759/1. The support of the funders is acknowledged. The views expressed in this piece are those of the authors and do not necessarily represent those of the funders.

Contact

You are also welcome to email us if you have any questions about this report or the work of the IRC generally: info@irc caucus.ac.uk

Cite as: Brown, A. February 2026. *An Integrated Approach to Estimating the Impact of UKRI*. Oxford, UK: Innovation and Research Caucus.

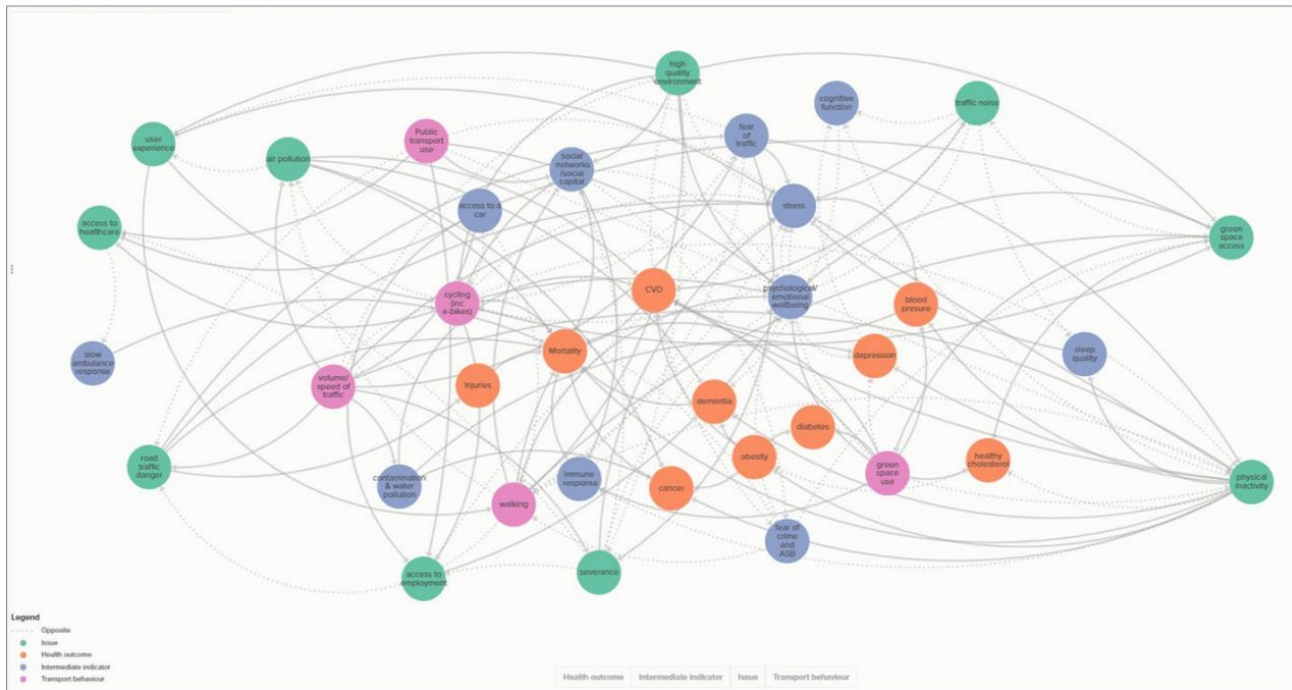
An Integrated Approach to Estimating the Impact of UKRI

Evaluating the impact of the activities of a large funding body such as UKRI involves assessing the effectiveness, efficiency, and broader outcomes of its investments across a wide range of programs, initiatives, and institutional support. As UKRI operates across multiple sectors, its activities have complex and long-term effects on systems, institutions, and society at large. The broader and more complex these activities are, the harder they are to capture robustly with a single evaluation; instead, individual activities or funding streams might be evaluated separately, and then their interaction effects carefully accounted for. This note discussed both the type of factors and impacts that might be investigated, and provides some overall suggestions as to the methodological approach that might be employed.

Theory of Change and Systems Mapping

The first step in any such complex impact evaluation is to develop a theory of change as to the separate impacts of the different activities and the interactions between them. The best current practice for this is to first develop systems maps, also known as causal loop diagrams. This is a visual depiction of the causal mechanisms through which the research activity funded by UKRI will affect the evolution of the UK economy. Theory of change frameworks have been applied in economic policy and evaluation contexts as a first step in clarifying causal assumptions and pathways (Corlet Walker, Mair & Druckman, 2018; Mason & Barnes, 2007). Developing such a theory helps ensure the causal chain from inputs through to outcomes is transparent and testable (Vun, 2021).

This should be done simultaneously at multiple levels. An overall systems diagram might show, at a high level, the different routes to impact resulting from the funding of research activity. Sub-maps might then show in more detail how the different activities—across councils, sectors, technologies, or geographies—function in their own right, and how these interact to create the overall effect. The different sub-maps need to be both internally coherent and comprehensive in their specific coverage, while also allowing for translation between maps and levels of the hierarchy. Precedents exist for using causal loop diagrams to study socio-economic systems, such as analysis of COVID-19 impacts (García-Olivares, 2021), agricultural market power (Biely, Mathijs & Van Passel, 2019), and the impact of transport infrastructure of public health outcomes (Cambridge Econometrics, 2021). A systems map from the latter project is shown below as an example.

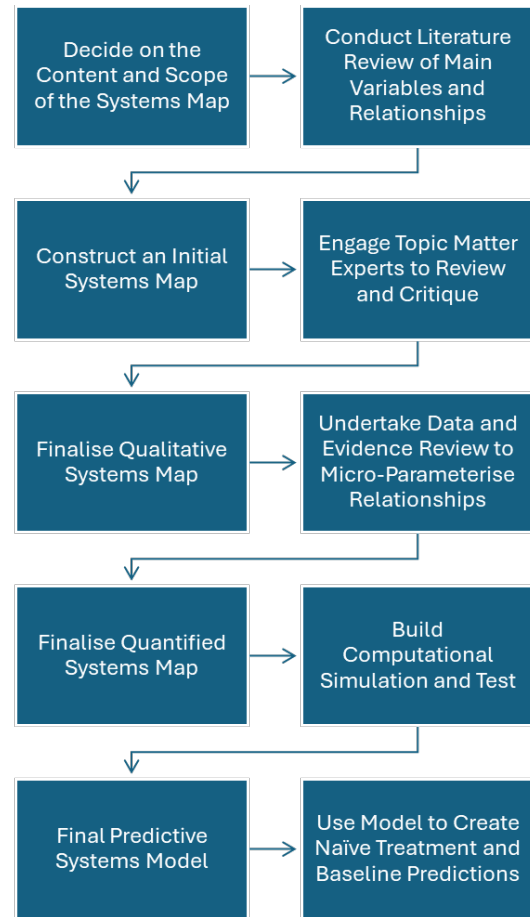


An example systems map based on literature review and expert engagement, detailed the linkages between transport infrastructure provision and population health outcomes. Key edges (causal relationships) in the diagram were also quantitatively estimated

Each map should include the main causal flow from inputs through to outputs and outcomes; other contributing factors that are either facilitating or confounding each of the different stages; any outcomes—positive or negative, intended outcomes or side effects—that result from the main causal flow or the way it interacts with contributory factors; and, where relevant, contraflow and feedback effects, including both direct node feedback loops and interactions with the master systems map and other sub-maps (see e.g., Groundstroem et al, 2021).

Map Development Process

The overall process through which the systems map might typically be developed would begin with initial framework development, where the segregation strategy is decided and the relationship between the master map and the sub-maps is determined. This would be followed by an initial literature and evidence review that would then feed into the initial development of each map and sub-map, showing the key causal relationships between the variables of interest. This should be supported by linked evidence for each node and each relationship within the map. Relational information should include details of the direction, magnitude, timing and spatiality of the anticipated effect.



The next stage would be expert engagement, identifying experts for each sub-map, who would be asked to comment on the flow of key causal relationships as they see them. Their input would allow the maps to be finalised in qualitative terms.

The final stage would be a quantitative evidence review, investigating the nature of the relationships between the individual nodes in each of the sub-maps. In some cases, this would use data available from the literature; in others, data may need to be constructed or calculated. In some cases, evidence may not be available and the relationship would need to remain qualitative only. Approaches linking causal graphical models and econometric estimation highlight how system-level structures can be formalised and empirically tested (Imbens, 2019; Klimek, Poledna & Thurner, 2019).

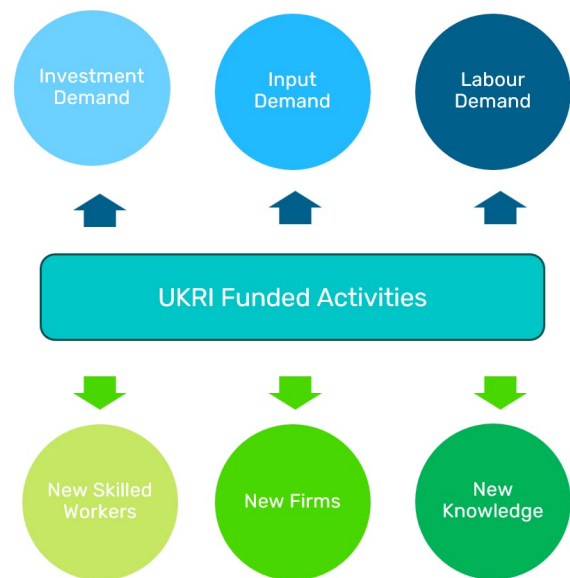
Systems Map content

The main causal flows that would need to be captured in any systems or sub-systems map would be the outcomes and the mechanisms through which the research activity funded by UKRI interacts with the economy, through both the stimulation of demand and the creation of supply. These should be estimated separately, but with interaction effects then accounted for.

On the **upstream** or demand side, UKRI-funded research activity could create demand for three

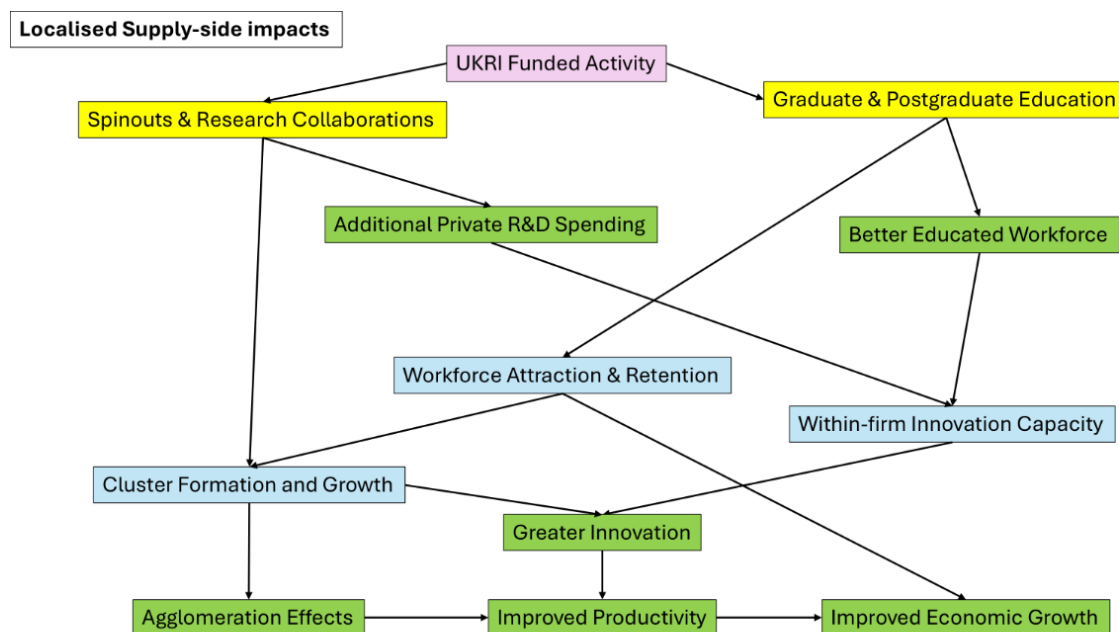
categories of research inputs: workers, capital equipment (such as machinery or property), and intermediate goods and services (a wide range from energy and material inputs, to administrative support and hospitality). However, we cannot assume that the existence of demand automatically leads to a one-for-one increase in economic activity. Instead, we must estimate how this demand would be met. In some cases, there may be existing excess supply, in which case the demand would automatically generate a one-for-one increase in output. In other cases, the demand may need to be met from outside the UK, or it may be met with activity displaced from elsewhere in the economy. A detailed estimation of substitution, leakage and displacement is typically required, and ideally this means considering the specific conditions of the local supply chain and labour markets; for example is this an area with under-deployed resources, or high levels of unemployment in general, or specific occupational underemployment that “matches” the occupational composition of the demand stimulus?

The impact of upstream effects tends to be easier to measure, to manifest over a shorter timeframe, and in a more localised geography, than downstream effects. In many cases, it is only the upstream effects that are quantitatively reported in appraisal and evaluation reports for this reason. However, many of the upstream effects are temporary and transient; as soon as the funding and activity is stopped, much of the effects are likely to disperse. Many of the longer-term effects of demand-side effects such as demand for skilled workers, specialist inputs, or equipment comes in the supply-side response to the demand stimulus, rather than the demand-effects directly. If knock-on investment in specialist equipment, skills and capabilities is stimulated, then these will still exist even after funding is withdrawn, providing



the possibility of alternative usages and the accumulation of local capabilities that form the foundation of a longer-lasting innovation cluster.

The majority of positive impacts will actually arise on the **downstream** or supply side, where UKRI-funded research activity creates a supply of skilled workers, new products or processes, directly commercialisable ideas (potentially leading to new businesses or entrepreneurial activity), and knowledge more generally. However, these are generally more complex to identify and more challenging to isolate, and are thus often omitted from appraisal and evaluation analyses. Again, we cannot assume that this increase in supply simply leads to a one-for-one increase in economic activity. We must consider whether there is existing demand and absorption capacity in the economy to make use of this increased supply, ask how the supply will be utilised, and consider secondary effects, spillovers, and feedback interactions. It is this element of the analyses where a fully-developed systems map is critical. Further examples of the kind of variables that might be included in a supply side analysis, and how they could potentially interact, are shown below.



A simplified systems map of different types of downstream economic impacts and their interaction points. The impacts considered could be much broader than those measured by economic metrics, and much more specific to a particular UKRI activity.

For a more systematic analysis of how research and development activity leads to economic value through the process of implementation, commercialisation, adoption and diffusion, our recent report [From research to productivity](#) provides a systems map that could be the starting point for this process. That report focused on generalised economic outcomes, such as labour productivity and consumer surplus; however, when constructing sector-specific system sub-maps, it would likely be appropriate to consider more detailed and relevant indicators, for example cancer remission rates and quality-adjusted life-years for a specific oncological investment return.

Hypotheses Development and Econometric Approaches

Once a detailed systems map and theory of change is developed and – potentially – modelled, the next key step would be to identify a series of hypotheses to be tested in the second stage of the analysis. These would be based on the contents of each of the sub-maps and chosen for their relevance and falsifiability. A combination of different types of hypotheses would be advisable, including both high-level hypotheses as to the overall impacts of UKRI-funded research activity on the UK economy, and specific, more easily testable components. For example, one hypothesis could be that UKRI activity might increase the development of a particular technology, followed by a second stage in which the development of that technology impacts a particular sector of the UK economy. In addition, specific examples of impacts could be used to demonstrate viability and to verify the broader set of hypotheses.

The standard approach to testing quantitative hypotheses is to collect the relevant historic real-world data and apply statistical methods to determine whether the hypothesis should be accepted or rejected. Within econometrics, one standard approach to doing this is the difference-in-differences (DiD) method (Card & Krueger, 1994; Callaway & Sant’Anna, 2021; Goodman-Bacon, 2021). This allows confounding variables to be controlled for in a systematic fashion, which is necessary because economics cannot replicate laboratory conditions (Angrist & Pischke, 2009). One best practice with DiD currently involves the construction of a *synthetic counterfactual* as a means of creating a control group against which the treatment group can be fairly compared (Abadie & Gardeazabal, 2003; Abadie, Diamond & Hainmueller, 2010). A synthetic counterfactual predicts what would have happened in the absence of intervention by constructing a weighted basket of comparators that best match the treatment case across a range of pre-intervention metrics.

Combining Systems Mapping and Econometric Approaches

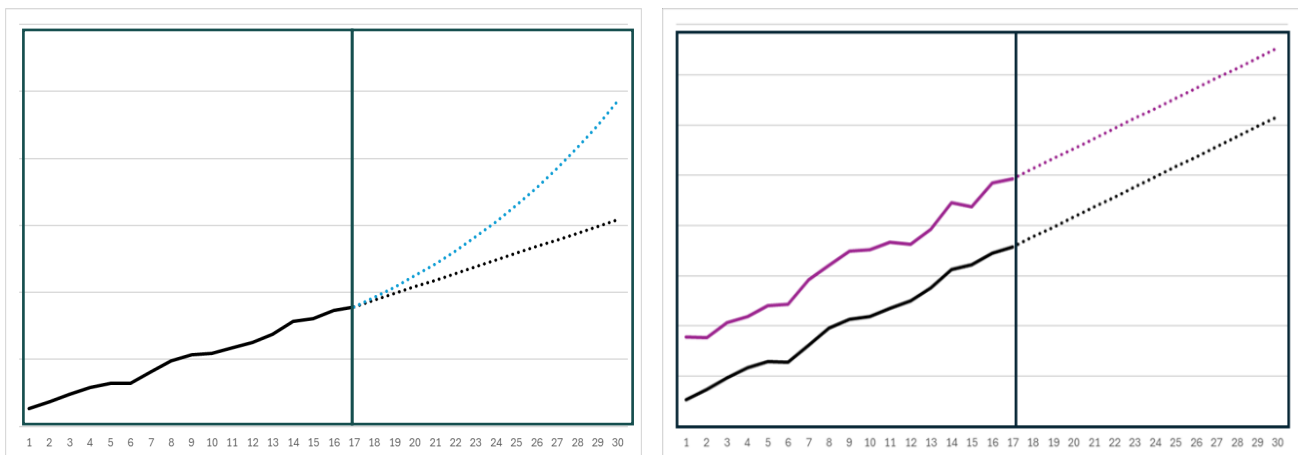
The purpose of the synthetic counterfactual is to internalise within the econometric analysis all of the potential factors that affect the metric of interest in the post-intervention period. In practice, this means constructing a credible estimate of the counterfactual outcome that would have occurred in the absence of the intervention, using information drawn from untreated units that display similar pre-intervention trajectories (Abadie, Diamond & Hainmueller, 2010). The underlying theoretical assumption is that external factors—macroeconomic shocks, policy changes, or sector-specific trends—affect both the treatment and control groups in the same direction and magnitude. If this holds, differences observed post-intervention can be attributed, with greater confidence, to the intervention itself rather than to confounding influences (Abadie, 2021).

It is therefore extremely important that the pre-intervention metrics chosen for constructing the synthetic control are selected with care. These metrics must adequately capture the relevant dimensions of economic structure and performance, including baseline output trends, input–output relationships, labour market dynamics, and sectoral demand or supply sensitivities (Cavallo et al., 2013). If important variables are omitted, or if poorly chosen predictors are used, the counterfactual may fail to replicate the underlying dynamics of the treatment unit, leading to biased inference.

In my view, the best way to strengthen the validity of this process is to use the quantified systems map to inform both the choice of metrics and the assessment of their adequacy. Systems mapping provides a structural representation of the interdependencies within the economy and the pathways through which research activity affects outcomes. By embedding this structure in a simulation framework, one can test whether the assumption of parallel external shocks is reasonable. If the simulated dynamics are sufficiently aligned, confidence in the robustness of the counterfactual increases (Firpo & Possebom, 2018; Imbens, 2019).

My recommendation, therefore, is to combine systems mapping with the synthetic counterfactual approach. The proposed sequence would be as follows. First, the quantified systems map is used to generate *ex ante* predictions for the treatment group's outcomes in both the impact and baseline cases, as if we were forecasting at the moment of intervention using only pre-intervention data. These initial forecasts can be termed the naïve impact case and the naïve baseline case. The units here are deliberately unspecified; they are intended to

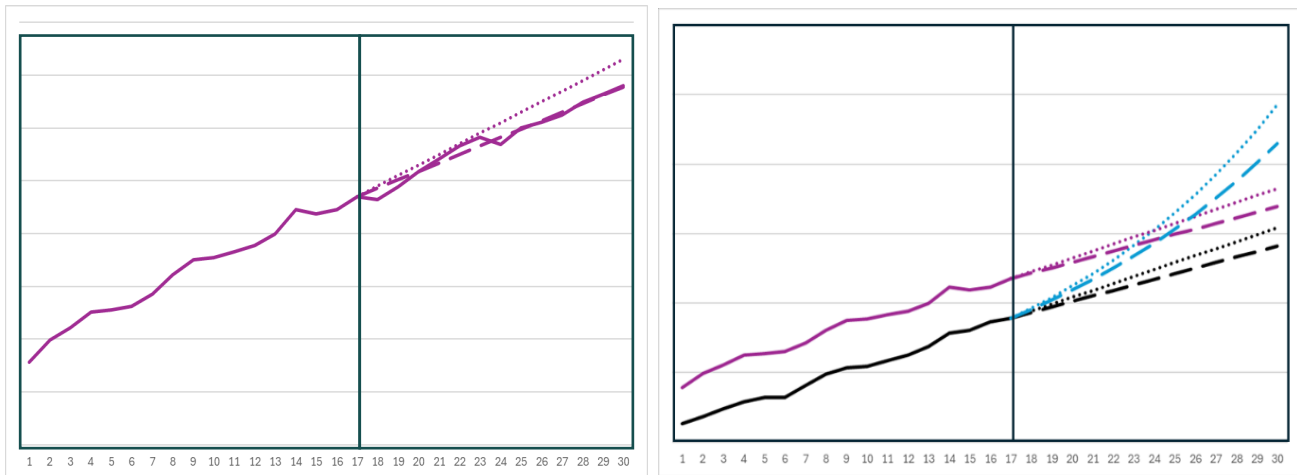
represent any one of a large number of economic, social or environmental outcomes, the outcome of which might be the basis of the hypothesis being tested.



On the left, the pre-intervention outturn leading to a naïve impact (blue dotted line) and baseline (black dotted line) projection for the post-intervention period, which would be generated by the systems model, for the treatment area. On the right, the selection of a synthetic counterfactual (purple line) based on alignment of both pre-intervention trends and post-intervention baseline

Second, these naïve predictions are used as benchmarks to guide the construction of the synthetic control group. Specifically, the synthetic control is assembled as a weighted average of non-treatment units that best replicate the treatment group's pre-intervention trajectory across the chosen predictors. The inclusion of systems-derived forecasts helps ensure that the predictor set is not only statistically relevant but also grounded in a theoretically coherent understanding of causal mechanisms.

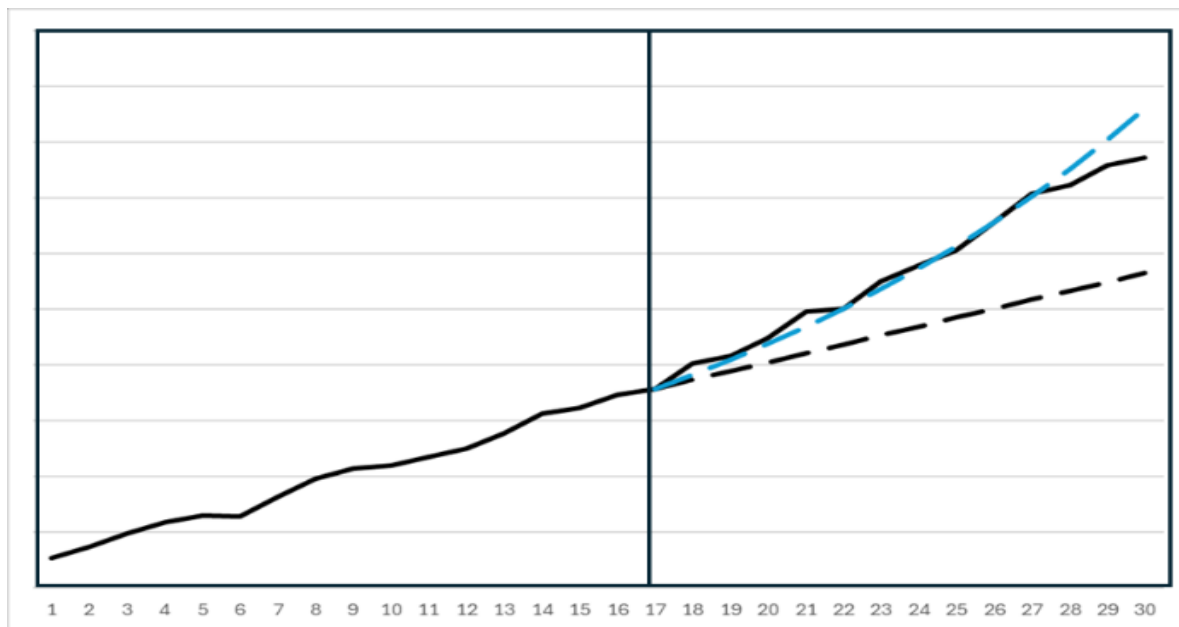
Third, once the synthetic control is established, the observed divergence between the treatment unit and its synthetic counterpart in the post-intervention period is used to adjust the naïve forecasts. This adjustment captures the contribution of omitted variables and unobserved shocks to the treatment dynamics. By applying the same adjustment to both the treatment and counterfactual predictions, we create adjusted treatment and adjusted counterfactual forecasts that incorporate both system-level causal structure and empirical evidence of unobserved heterogeneity.



On the left, the comparison of the synthetic counterfactual baseline projection to post-intervention outcome, and subsequent adjustment of the baseline projection (purple dashed line). On the right, applying the same adjustment to the impact and baseline projections for the treatment case (blue and black dashed lines).

Finally, the observed outcome is compared against these adjusted forecasts. The hypothesis test then reduces to assessing whether the actual outcome is statistically closer to the adjusted counterfactual prediction (suggesting no impact of the intervention) or to the adjusted treatment prediction (suggesting a significant causal effect). A goodness-of-fit test or Wilcoxon signed-rank test could provide an appropriate measure of statistical significance.

This hybrid approach combines the interpretability and theoretical coherence of systems mapping with the statistical rigor of synthetic control estimation, thereby improving both internal validity and external interpretability of causal claims (Arkhangelsky et al., 2021; Doudchenko & Imbens, 2016)



Comparing the post-intervention treatment outcome (solid black line) to the adjusted impact and baseline projections. In this theoretical example, it's likely that the alternate hypothesis would be accepted and a significant causal relationship would be identified

Now that you have read our report, we would love to know if our research has provided you with new insights, improved your processes, or inspired innovative solutions.

Please let us know how our research is making a difference by completing our short feedback form [via this link](#).

You are also welcome to email us if you have any questions about this report or the work of the IRC generally: info@ircaucus.ac.uk

Thank you

The Innovation & Research Caucus

References

- Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque Country. *American Economic Review*, 93(1), 113–132.
- Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, 105(490), 493–505.
- Angrist, J. D., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G., & Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review*, 111(12), 4088–4118.
- Biely, K., Mathijs, E., & Van Passel, S. (2019). Causal loop diagrams to systematically analyze market power in the Belgian sugar value chain. *AIMS Agriculture and Food*, 4(3), 711–730.
- Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2), 200–230.
- Card, D., & Krueger, A. B. (1994). Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review*, 84(4), 772–793.
- Corlet Walker, C., Mair, S., & Druckman, A. (2018). *A Theory of Change Approach for Measuring Economic Welfare Beyond GDP*. CUSP Working Paper No. 10. Centre for the Understanding of Sustainable Prosperity.
- Doudchenko, N., & Imbens, G. W. (2016). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. NBER Working Paper No. 22791.
- Firpo, S., & Possebom, V. (2018). Synthetic control method: Inference, sensitivity analysis and confidence sets. *Journal of Causal Inference*, 6(2).
- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2), 254–277.
- Groundstroem, F., Juhola, S. Using systems thinking and causal loop diagrams to identify cascading climate change impacts on bioenergy supply systems. *Mitig Adapt Strateg Glob Change* 26, 29 (2021). <https://doi.org/10.1007/s11027-021-09967-0>
- Imbens, G. W. (2019). Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics. arXiv:1907.07271.
- Klimek, P., Poledna, S., & Thurner, S. (2019). Economic resilience from input–output susceptibility improves predictions of economic growth and recovery. arXiv:1903.03203.
- Mason, P., & Barnes, M. (2007). Constructing theories of change: Methods and sources. *Evaluation*, 13(2), 151–170.
- Vun, G. (2021). Analysing the credibility of theories of change. In *Theories of Change in Sustainable Finance* (pp. 39–63). Springer.

Appendix

- » How do we ensure an approach that considers less “reportable” investments, including infrastructure, institutes, and QR?

The philosophy of dynamic systems modelling is one of exploration and understanding, rather than point-estimate prediction. Whereas in some more linear modelling approaches, more uncertain or qualitative variables might be consciously excluded from the analysis, the systems approach is to try to include everything that the evidence suggests is important in order to understand it better, even if they have large error bars around any possible quantification estimates. The systems analyst would then undertake sensitivity testing to try to understand how these important but challenging variables might interact with the rest of the system. In cases where outcome variables are difficult to quantify numerically, specific hypotheses may instead be developed to be tested qualitatively, for example through systematic interviews, rather than through econometric methods.

- » How do we ensure that we capture UKRI’s full value, including non-funding roles such as convening, stewardship, and system leadership?

The cornerstone of this approach is that if you can accurately, coherently, and with supporting evidence, articulate what the direct effect of UKRI activities such as convening and stewardship, might be - for example with the use of example case studies - then they can and should be included as qualitative variables in a systems map, where they can be included as factors in both the hypothesis formation and result interpretation stages of the analysis wherever relevant. As above, they can either then be quantified - but with large error bars and appropriate sensitivity testing - or they can be explored with more qualitative approaches.

- » For any proposed ROI figure, how should monetisation or quantification be handled without over-claiming?

Any final figure should be reported as a range, rather than as a point estimate. This should include both standard error bars, and where appropriate the effects of varying different fundamental assumptions or excluding specific “hard-to-quantify” mechanisms or effects.

- » If ROI is not the most appropriate headline metric, is there an alternative that still provides a clear and compelling summary of UKRI’s contribution?

The short-term and long-term costs and benefits of both individual and combined UKRI activities and workstreams are worth understanding in detail. Two important points: it's important to note that in economic systems, whether something is a cost or a benefit is often contextual – for example increasing labour demand is beneficial in a time of recession or stagnation, but less so in times of inflation and labour shortages; and related to this; all impacts should be reported against a clear and realistic counterfactual: constructing and articulating such a counterfactual is often the most challenging part of any impact analysis.

Author-specific questions:

- » Should a single systems map be developed for all UKRI impact pathways, or separate maps by activity? How can complexity be managed? Given UKRI's scale, what high-level time and FTE requirements might this involve?

The answer to many of these questions would become clearer in the early stages of any analysis. Clearly, a high-level overarching systems map would need to be produced that maps all UKRI activities and their inputs, outcomes and key interaction points, as well as at least one sub-system map (maybe several more) that map out in more fine-grained detail the activities of a specific programme, council, facility or workstream, that might look at the knowledge that has been generated, the technologies that have been developed, the workers attracted or developed, the firms supported, and try to understand and quantify how the sectors have subsequently evolved differently.

What is difficult to predict is the extent to which one template sub-systems map could be generated that could then be modified for different programmes, or whether the impacts of different programmes are so qualitatively different that this would not be possible.

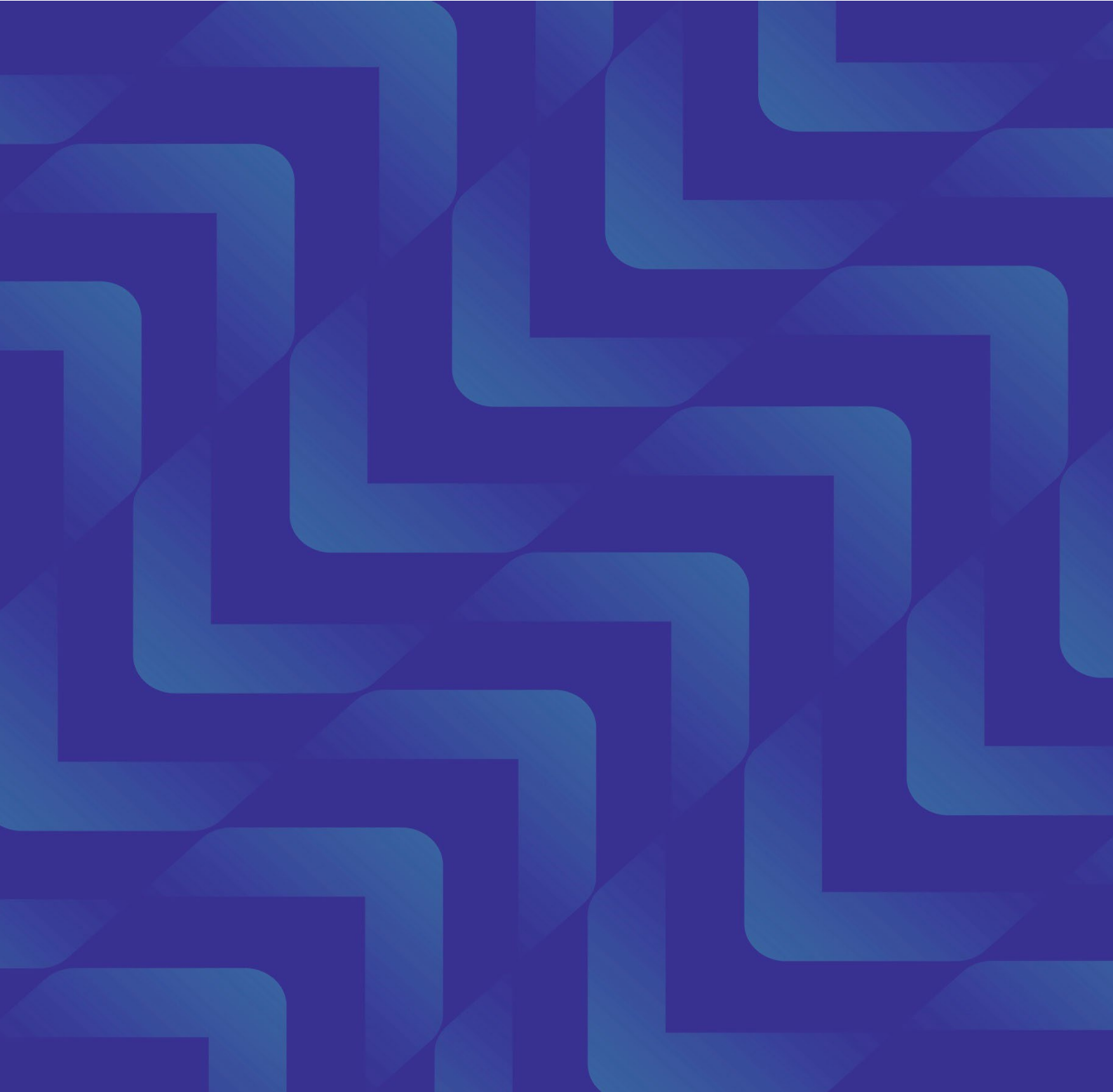
In terms of duration and resources, somewhere between 9 months and 120 FTE days, to 18 months and 240 FTE days, depending on complexity and depth.

- » How does a systems map translate into econometrics? Would every pathway be tested, or only a subset? How should pathways be prioritised for synthetic control analysis?

There are two stages to this – firstly the systems map needs to be parameterised as fully as possible in order to become a systems model. Parameters are often available in the literature, or can be measured directly from data, others will be approximated and sensitivity tested.

This system of parameterisation allows us to understand the dynamics of the system as a whole, and to identify the dominant and modifying pathways, based on the extent to which they determine the dynamics of those pathways determine the behaviour of the system as a whole.

It is these dominant pathways that are more likely to be used as the basis for a quantifiable hypothesis. For example, “increasing input X led directly to a change in output Y”: there are probably many causal pathways through which X influences Y, but if our initial modelling tells us that for the range of realistic parameters, some pathways have a much larger effect than others, and therefore we should focus the econometric analysis on these.



www.ircaucus.ac.uk

Email info@ircaucus.co.uk Twitter [@IRCaucus](https://twitter.com/IRCaucus)



Delivered with
ESRC and
Innovate UK