Department for Business, Energy & Industrial Strategy

EVALUATION OF POLICIES FOR LOCAL ECONOMIC GROWTH: SCOPING STUDY

BEIS Research Paper Number 5

OCTOBER 2017

Contents

Introduction to the study

The report by the National Audit Office (NAO), 'Evaluation in Government'[1] , is critical of the quality of ex post evaluation and the attribution of impacts to policy interventions. Standards of evaluation quality, and the ability to explain attributed impacts using a robust counterfactual, were considered variable. The evaluation of business and spatial interventions was considered to be generally weaker than evaluations covering education and the labour market.

In this context, the aim of this study was to establish the scope to use methodological approaches for use in ex post evaluation. These approaches should be capable of providing a more rigorous assessment of the extent to which interventions have had an impact on local economic development outcomes that would otherwise not have occurred in the absence of the programme. This in turn requires methods capable of providing a more formal comparison of actual outcomes with a statistically robust counterfactual reflecting the likely outcomes in the absence of the interventions concerned.

The study does not seek to scope methods that would assess whether the interventions had achieved significantly different outcomes and/ or value added compared to any alternative programme that might be defined.

More specifically, the scoping study is required to:

- identify a set of methodological options and recommendations for undertaking impact evaluation of interventions – segmented as appropriate for different types of intervention;
- provide an indication of the appropriate timelines for, and financial costs of, implementing the recommended evaluation options;
- describe the recommended methodological approaches in a technical annex and present it in sufficient detail so that it can be understood by any selected evaluation contractor at the full evaluation stage; and
- identify any immediate actions required to facilitate the impact evaluation, including the collection of baseline or monitoring data.

---

[1] Evaluation in Government, Report by the National Audit Office, December 2013

Monitoring and evaluation requirements

General principles

The general principles for dealing with resources used by public sector organisations in the UK are set out in Managing Public Money[2]. In particular, it explains the importance of evaluating public interventions, and emphasises that Parliament expects accounting officers to take 'personal' responsibility for 'ensuring that the organisation's procurement, projects and processes are systematically evaluated and assessed'. Similarly, the Civil Service Reform Plan[3] explains that accounting officers 'must be accountable for the quality of the policy advice in their department'.

Recently however, the evaluation of activities funded by public sector organisations has been increasingly criticised for its deficiencies. This prompted the NAO to undertake a comprehensive assessment of impact and cost-effectiveness evaluations across the 17 main departments and some of their bodies[4]. Specifically, the report provided a quantitative assessment of the coverage, quality and use of evaluation evidence, as well as the resources spent on evaluating the public sector activities. The main findings of the report can be summarised as follows:

- There are significant gaps in the coverage of evaluation evidence.

- The existing evaluations are of varying quality, with spatial policy and business support evaluations generally weaker than those covering labour market and education policies.

- Evaluation reports that are weaker in identifying causality tend to be more positive in assessing what the intervention achieved.

- There are limited references to past evaluation evidence within the evaluation reports reviewed.

- Independent researchers face difficulties in accessing administrative data and other government data to conduct their own evaluations of government interventions.

- The lack of demand from policy colleagues and the tight evaluation timescales represent significant barriers to the production and use of evaluation evidence.

The approaches explored in subsequent sections are designed to address such weaknesses in particular through the use of comparison group based methods designed to assess the likely relevant economic outcomes in the absence of the different types of investment (i.e. the counterfactual).

---

[2] HM Treasury, July 2013, Managing Public Money
[3] HM Government, June 2012, Civil Service Reform Plan
[4] NAO, December 2013, Evaluation in government;
http://www.nao.org.uk/wp-content/uploads/2013/12/10331-001-Evaluation-in-government_NEW.pdf

Potential Modelling Approaches

Introduction

This section introduces some of the concepts that underpin the technical elements of the scoping exercise that follows. It provides a narrative explanation of the underlying issues and methods that are of prominence in programme evaluation. It does not address all such considerations but concentrates on the key features that will guide understanding of the approaches that are reviewed and. Both the discussion in this section, and the 'Estimating Gross Impacts' sections, operate on the basis that due account is taken of potentially contaminating policy interventions through the modelling process.

Improving the strength of evaluation analysis

Policy evaluation is straightforward in conceptual terms. The process starts with an intervention that is designed to i) directly affect or assist individuals or firms in some way (e.g. employment support, business advice) ii) indirectly affect or assist individuals or firms through some means (e.g. improve transport links or roll-out of broadband). The evaluation question is simply whether or not the policy works as intended and what, if any, the scale of impact is.

Unfortunately, answering the question is not so straightforward. One issue is that it is impossible to simultaneously observe what happens to any particular individual or firm. They are either 'treated' (and receive assistance), or 'not treated', and cannot be in both states at the same point in time[5]. Another issue is whether we can be sure that any identified difference or impact is due to the policy intervention as opposed to some other influence. It is addressing these issues that makes programme evaluation complex.

In terms of being unable to observe the treated/non-treated states of any individual or firm, one way around this is look at what happens to a cohort of both. In this way one can compare the average performance of the treated (after the intervention) to the average performance of those untreated (after the intervention).

It would be easy to jump to the conclusion that any difference is due to the policy intervention and that we thereby have an estimate of the average causal effect (ACE) of the policy[6]. Some care is, however, required before coming to this conclusion as it only really applies if one is comparing like with like. If the treated and non-treated cohorts are exactly alike in all observable and unobservable characteristics (i.e. effectively identical) then one would expect:

- the non-treated group to behave in the same way as the treated group if the latter had not been treated; and
- both the treated and non-treated groups to behave in the same way in response to any influences beyond the policy intervention.

---

[5] A situation often referred to as the 'fundamental problem of causal inference'.
[6] As opposed to the individual causal effect (ICE) that we cannot calculate.
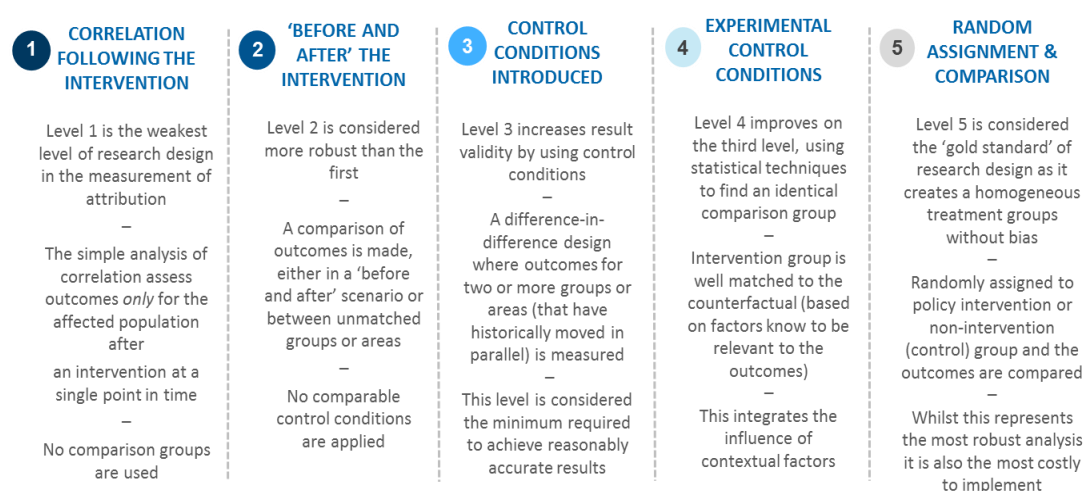
In these circumstances:

- the non-treated cohort provides a perfect counterfactual and represents a perfect control group;
- since both groups react in the same way to observable and unobservable influences, any deviation in performance post-intervention can be attributed to the intervention; and
- estimates of policy impact will be unbiased.

This state of affairs – a treatment and control group that is perfectly balanced in terms of observable and unobservable characteristics – represents an ideal policy evaluation environment and provides something of a benchmark in terms of experimental design with robustness of the latter gauged by how close the design comes to replicating this set of conditions. In essence:

- a strong design will be able to define treatment and control groups that are equivalent in terms of both observable and unobservable characteristics;
- a moderately strong design may not have the desired level of equivalence but may be able to moderate the extent of any subsequent bias through application of statistical methods; and
- a weak design will not be capable of moderating any bias due to non-equivalence and may not even use a treatment/control group structure.

Something akin to this perspective is contained in the Maryland Scientific Methods Scale (SMS) – used by the researchers investigating public policy interventions on behalf of the National Audit Office[7] and as the basis of the evidence reviews undertaken by the What Works Centre for Local Economic Growth – which formalises the robustness of evaluation design along a scalar from weak (level 1) to strong – level 5 (see Figure 1 for a description).

Figure 1: The Maryland Scientific Methods Scale

1 CORRELATION FOLLOWING THE INTERVENTION	2 'BEFORE AND AFTER' THE INTERVENTION	3 CONTROL CONDITIONS INTRODUCED	4 EXPERIMENTAL CONTROL CONDITIONS	5 RANDOM ASSIGNMENT & COMPARISON
Level 1 is the weakest level of research design in the measurement of attribution – The simple analysis of correlation assess outcomes *only* for the affected population after an intervention at a single point in time – No comparison groups are used	Level 2 is considered more robust than the first – A comparison of outcomes is made, either in a 'before and after' scenario or between unmatched groups or areas – No comparable control conditions are applied	Level 3 increases result validity by using control conditions – A difference-in-difference design where outcomes for two or more groups or areas (that have historically moved in parallel) is measured – This level is considered the minimum required to achieve reasonably accurate results	Level 4 improves on the third level, using statistical techniques to find an identical comparison group – Intervention group is well matched to the counterfactual (based on factors know to be relevant to the outcomes) – This integrates the influence of contextual factors	Level 5 is considered the 'gold standard' of research design as it creates a homogeneous treatment groups without bias – Randomly assigned to policy intervention or non-intervention (control) group and the outcomes are compared – Whilst this represents the most robust analysis it is also the most costly to implement

Source: Adapted from Sherman et al. (1997)

---

[7] National Audit Office (2013), Evaluation in Government

The SMS forms the prism through which our analysis of potential evaluation methods is undertaken in subsequent sections and it is helpful to walk through the basis on which higher scores (more effective design) are assigned.

Figure 2 indicates the broad options available for the evaluation design and summarises the discussion of methods and approaches that follows in subsequent sections.

Figure 2: Evaluation design options using statistical modelling approaches
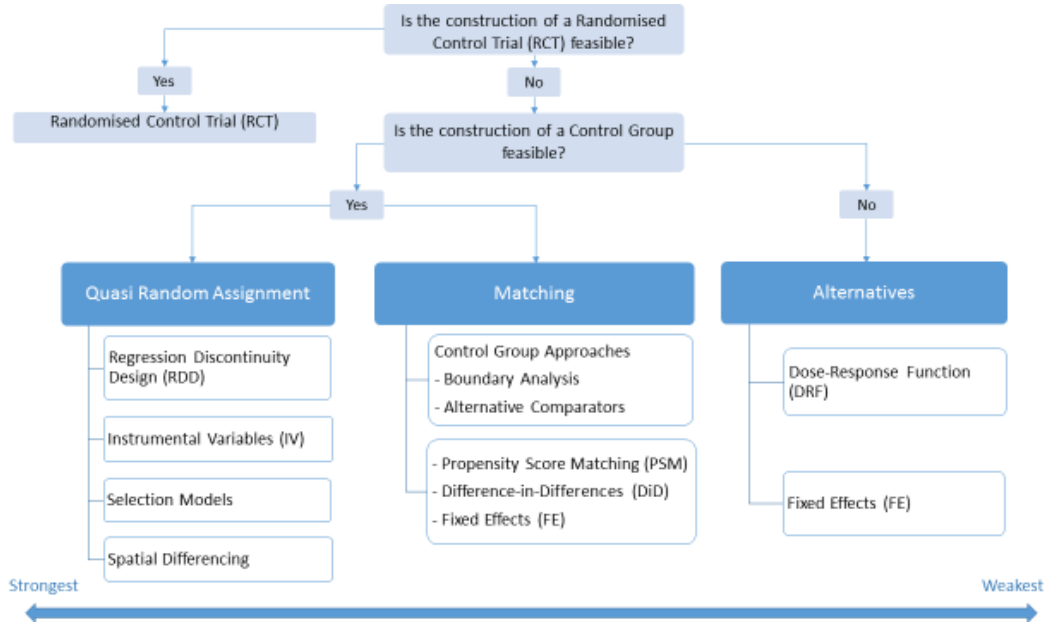
Experimental Methods: Random Assignment

The highest SMS score (level 5) is reserved for an evaluation design that comes very close to the 'ideal' discussed earlier and typically takes the form of a random control trial (RCT).

The words 'random' and 'control' are the key elements to this approach. The presence of a control group is a pre-requisite in an attempt to replicate the condition of the 'ideal' benchmark outlined above but the concept of randomisation (or random assignment) is also crucial.

Unlike the benchmark scenario, where identical treatment and control groups can simply be assumed for purposes of exposition, constructing comparable groups is likely to be a challenge in practice. It is not difficult to see why this is the case because no matter how much effort is invested in ensuring that members of the groups are equivalent (or even identical) in terms of observable characteristics, there is the problem of unobservable characteristics.

These characteristics cannot be observed and therefore cannot be measured but if they vary significantly across subjects (individuals, firms) then comparability between treatment and control groups may be compromised.

It may be the case, for example, that some unobservables encourage certain types of subjects to 'select into treatment'. Since these features cannot be identified, it is difficult to define a control group that is equivalent.

It may also be that policymakers select participants in programmes that they 'feel' will respond more effectively to the intervention. Once again this makes the definition of an appropriate control group difficult.

In the face of such 'selection bias', any assessment of policy impact through comparison of groups may be inaccurate or biased. This unobservable heterogeneity means that we can no longer claim that the control group will behave in the same way as the treated group would have behaved if the latter had not been treated. Part of what we might assume to be a policy impact might simply reflect lack of comparability between the treatment and control group due to differences in unobservables. The true policy impact might be higher or lower than that reported.

The reason that RCTs are scored highly in the SMS scale is that they offer a way of dealing with these issues:

- prior to any intervention, substantial effort is made to ensure that all members of potential treatment and control groups are statistically equivalent in terms of observable and (where feasible) unobservable characteristics[8];
- membership of each group is determined by random assignment.

---

[8] The latter might be achieved by some process designed to elicit information on unobservables which can then be used to improve equivalence between the groups.

In this way composition of groups is as equivalent as it is possible to be. Random assignment will ensure that the distribution of unobservables will broadly be the same between the groups and there is no issue of selection bias.

There are a variety of circumstances and interventions in which pure experiments of this nature are feasible though substantial care is required in design and there are a number of potential issues which can complicate implementation:

- attrition effects whereby the size of groups declines as individuals drop out from the experiment (less of an issue where the latter is random and affects both groups in a similar way);
- non-compliance whereby subjects move between groups;
- randomisation bias whereby the randomisation process itself results in a situation where a substantial proportion of an eligible population are not represented in the experiment;
- Hawthorne effects where subjects respond differently in the light of taking part in an experiment; and
- substitution effects in which control group subjects seek alternatives in the light of not being selected for treatment.

Above and beyond technical issues there are (on occasion) considerations regarding the ethics of randomisation, the feasibility of constructing an adequate control group in a wide-ranging all-embracing intervention and a question as to whether it is possible to ensure an assignment is truly random and not subject to the preferences/desires of administrators to target a particular cohort or improve policy results.

Recent examples of the use of RCTs can be found in the innovation policy field. For example, Van der Steeg et al. (2006) examines an innovation voucher scheme which provides SMEs the opportunity to engage with research institutions and where the selection of SMEs is based on a lottery. There has also been recent efforts made, in particular by NESTA, with the establishment of Innovation Growth Lab (IGL), which looks at piloting the use of randomised control trials in innovation policy and looks at establishing a database of RCT trials in innovation and entrepreneurship.

However, as far as local growth policy interventions are concerned, we do not envisage that RCTs will play a major role in the evaluation process. There are some interventions that might prove suitable to an RCT approach but they are complex to organise effectively and costly to implement.

Quasi-Experimental Methods: Quasi-Random Assignment

There are many instances in which random assignment experiments are not feasible. In this context, the next best thing may be an approach whereby something close to a random assignment has (possibly inadvertently) been used (or can be used) to generate treatment and control groups – the groups are structured 'as if' they were randomly generated and are therefore very similar in terms of observable and unobservable characteristics. The two approaches most often employed in this context are regression discontinuity design (RDD) and instrumental variables (IVs); both are generally scored at level 4 on the SMS scale.

Regression Discontinuity Design (RDD)

A potential opportunity for a quasi-random approach occurs if treatment and control groups are defined according to a set of criteria or rules that generate a discontinuity in the probability of treatment at a given threshold. It may be, for example, that subjects wishing to participate in a project are scored on some basis with a cut-off for receiving support determined by the level of resource constraint available.

In this case, the threshold 'happens' to be placed where it is on the basis of a resource constraint that will have been determined well before the programme of support commences. It is a 'relatively random' event and the profile of observable and unobservable characteristics is likely to be similar around the threshold[9].

More generally, the principle for RDD is that there is an assignment (or forcing) variable with a cut-off above/ below which treatment does/ does not take place. It is then argued that this process is quasi-random in nature and that subjects just below the cut-off provide a good control group for those just above. For example, a business or innovation support scheme may allocate a point score to all firms applying and only support those that meet a certain threshold (say 75 out of 100 points). In this scenario, a firm that score 76 points will only be marginally better than the one that scores 74 points, but the consequences of the scores will be very different. The firm with the 76 point score will receive support whilst the other firm will not be supported. Comparing the outcomes for these two firms is thus indicative of the causal impact of the support.

Two variants of the approach are typically referenced to reflect situations whereby there exists strict adherence to the cut-off threshold - there is a sharp discontinuity with no subject crossover and all those selected receive treatment (no 'no-shows') – and situations where the rules are not applied so strictly, where some subjects cross over or there exist some no-shows – the so called fuzzy discontinuity.

The RDD approach, in both its Sharp (SRD) and Fuzzy (FRD) guises, has become a popular method of programme evaluation. As with RCTs, however, there are some issues that require careful attention. In particular, RDD methods can be compromised if:

- subjects can (precisely) manipulate the assignment/forcing variable, which would negate the quasi-random nature of the process;

---

[9] This is less as one moves further away from the threshold.

- the cut-off threshold is not determined independently of the assignment variable;
- there are discontinuities in other characteristics – that is there is no other basis for treating subjects on either side of the threshold differently.

As with instrumental variables, the potential to employ RDD approaches is examined within the scoping exercise.

Instrumental variables

One way of approximating random assignment is through the use of instrumental variables (IVs). Instrumental variable techniques have a long pedigree in statistics and have proved useful in a number of circumstances.

As noted earlier, in the absence of an RCT, and the equivalence of treatment and control groups, any change in the performance variable being used to assess policy impact may reflect differences between the assisted/non-assisted groups as well as the intervention.

IV approaches provide a way of addressing this problem and do so by seeking to work out what would the performance profile look like if there was a way of ensuring that treatment and control groups were more alike.

They do this by constructing an 'instrument' which is highly correlated with treatment status (treated/not-treated) but not correlated in any way with the performance variable itself. In other words, the instrument is used to predict which individuals/firms should be regarded as treated/non-treated and the treatment and control groups are reconstructed on this basis. Performance is assessed in relation to these reconstructed groups rather than the original groups.

It is clear that the process only works effectively if the instrument is a credible substitute for the variable it replaces is:

- strongly correlated with participation status;
- completely exogenous (not affected by the policy outcome); and
- excludable so that there is no direct impact on the policy outcome.

In practice, the identification of an appropriate instrument can prove challenging. Some recent innovative uses of instrumental variables includes Garcia-Lopez et al (2015) which uses Roman roads and 1760 Bourbon road profiles as instrumental variables to investigate the role of transport infrastructure in encouraging suburbanisation across Spain. Similarly, Holl (2014) uses the 1760 postal routes to examine the link between highways and productivity in urban and rural Spain.

Instrumental variables have been used widely in a variety of circumstances and they are considered as an option in examining evaluation of interventions.

Selection models

Selection models are a variant of the IV approach. These seek to directly define the basis on which a subject selects/is selected into treatment. As long as it is possible to account for participation in a strong and robust manner, then this information can be used as a

basis for offsetting any selection bias in contrasting performance between a treatment and control group. The quality of the process is clearly dependent on the extent to which the 'selection' first-stage mimics the generation of a strong instrumental variable as described above and attains the same characteristics.

Spatial differencing

Another recent and emerging technique is that of spatial differencing. In contexts where policy intervention is spatially bounded, or targeted at particular geographies, such boundaries can act as form of discontinuity. If the unobservable characteristics of neighbouring areas vary smoothly or continuously across the boundary at which policy eligibility/intervention ceases, then it is possible to use (policy) treated and non-treated neighbouring areas as a basis for impact assessment. This developing method is considered an option in relation to interventions where intervention may involve definition of boundaries.

Quasi-Experimental Methods: Matching

The primary benefit of random and quasi-random approaches to evaluation is that they provide a credible basis for generating treatment and control groups that match in terms of observable and unobservable characteristics, providing a robust basis for assessment of policy impact.

Nevertheless, there exist circumstances in which such approaches are not viable. Scenarios in which an assessment of policy impact is required, where a policy intervention has already occurred on a non-random basis and where no quasi-random approach is feasible, are relatively common.

Here, there is little alternative but to pursue quasi-experimental methods that seek to mimic an experimental approach in terms of contrasting treatment and control groups while making every effort to offset the influence of any differences in observable or unobservable characteristics that may serve to undermine the credibility of policy impact conclusions.

Much of the focus in quasi-experimental methods is thereby on the construction of control groups and addressing the time varying/ invariant nature of observable/ unobservable characteristics that are of less concern in random assignment approaches.

Propensity score matching

One of the most common mechanisms for constructing control groups is Propensity Score Matching (PSM).

The objective here is to identify treatment and control groups that are very close in terms of observable characteristics. Instead of trying to match on the basis of individual

characteristics, however, the match is made on the likelihood or propensity (effectively the probability) to participate or be treated[10].

While simple in concept, there are some demanding aspects to the technique including a sufficient number of subjects that (despite being treated or non-treated) display similar characteristics profiles (the common support requirement). It is also necessary to address a range of considerations such as how close a match is acceptable and should every treated subject have one or more than one match.

The main drawback with PSM is that it can only match on the basis of observable characteristics and does not address the issue of unobservable characteristics. As such, and by itself, matching is scored at level 2 on the SMS scale. Consequently, PSM is often combined with other techniques of which difference-in-differences (DiD) analysis is particularly prominent.

Difference-in-Differences (DiD)

Difference-in-differences techniques are based on a simple contrast between the (before and after intervention) difference in outcomes for both a treatment and control group.

The approach improves on matching because, in introducing a time element, it provides a way of dealing with unobservable differences between treatment and control groups.

Introducing time allows one to distinguish between time-invariant and time-variant unobservables. This is useful because if there are some unobservables that are time invariant, then any change in performance over time cannot be due to these unobservables as they do not change with time.

Consequently, on the assumption that treatment and control groups have already been matched, and that the composition of the groups does not alter, the primary benefit of the technique is that the before/after contrast can account or control for time-invariant unobserved characteristics between the treatment and control group. As such, matching and DiD, combined, achieves a level 3 SMS score. The approach, does not, however, account for unobservable differences that are time-varying.

To operate effectively, one of the key requirements of a DiD approach is that both the treatment and control groups would follow the same time-trend in the absence of intervention. Difficult to verify, most approaches examine pre-intervention time trends between the two groups to examine evidence of parallel trends.

Combined matching/DiD approaches are widespread in programme evaluation studies and are considered along with other approaches in the scoping exercise.

Fixed Effects (FE)

In some situations it is possible that a longitudinal dataset of observations are available for both treatment and control group members. Fixed effect (FE) panel data methods are

---

[10] The use of generalised propensity scores for continuous policy measures is discussed later.

available for this scenario and, given information for prior/post intervention, can account for observed/unobserved, time-invariant differences in characteristics. Similar in many ways to the DiD format, a combined matched/FE approach will generally score at level 3 on the SMS scale. Combined matching/FE approaches are considered along with other approaches in the scoping exercise.

Control Group Variants

As matching approaches are extended to more diverse interventions, it is not uncommon for the nature of control groups to be defined in different ways. Later sections of this report reference three types of group structures:

Control group approach:

- this is the standard control group approach in which a set of validated controls are constructed to be contrasted with a treatment group;

Control group – boundary approach:

- the definition of control groups immediately beyond the 'boundary' of interventions is an increasingly common modus-operandi. The rationale is that any variance in observables can be adjusted for in modelling, while a control group located in physically proximate areas to the intervention area means that area or economy-based unobservables are likely to affect both the treatment and control groups in a similar manner[11];

Control group – alternative comparators:

- alternative comparator approaches are another mechanism that seeks to address potential differences in unobservables between treatment and control groups. Comparators may, for example take the form of subjects treated late in the intervention regime. If it is possible to assume that the later subjects, in seeking and receiving assistance, are likely to have similar unobservable characteristics to previous beneficiaries, then they may form an effective control group for earlier beneficiaries.

Alternative Methods

Dose-Response models

In most of the methods outlined above, treatment and non-treatment are represented as a binary measure (1/0). In practice, some forms of treatment might also be represented as a continuous measure (perhaps amount of spend) and policymakers might be interested in assessing intensity of spend.

---

[11] This approach is not the same as Spatial differencing discussed in the previous sub-section where boundary definition is, or can be argued to be, quasi-random in nature.

This provides an opportunity to consider dose-response models which relate level of performance to level of treatment with non-assistance represented by a zero treatment level. Recent development of generalised propensity score (GPS) methods, operating in the same way as PSM with similar considerations regarding common support and balancing, provides a basis for modelling the dose-response function (DRF) to examine differential impact profiles.

Summary and conclusions

There is a growing body of work testing and applying quasi-experimental methods as the basis for evaluation and impact assessments. These methods attempt to mimic as closely as possible the 'gold standard' of a Random Control Trial (RCT) approach in contexts where this is not a viable option.

The available methods have varying potential information requirements and limitations; they vary in terms of their likely statistical robustness and none is capable of application in all contexts. Subsequent sections assess the suitability of the available methods in the assessment of the different intervention types. Use of different approaches for different types of intervention will be appropriate; and the presumption is that it will be possible to assess at least some of the major gross economic impacts of a significant proportion of the projects funded under almost all of the intervention types using the various methods discussed above.

We assess the feasibility of these approaches to each type of intervention in turn, identifying those approaches that seem most appropriate and setting out the basic steps for the application of the preferred approach. For each type of intervention we start by summarising the nature of the intervention (based on the information provided) and the relevant indicators as the basis of subsequently assessing feasible approaches to statistical modelling of the gross impacts of the intervention.
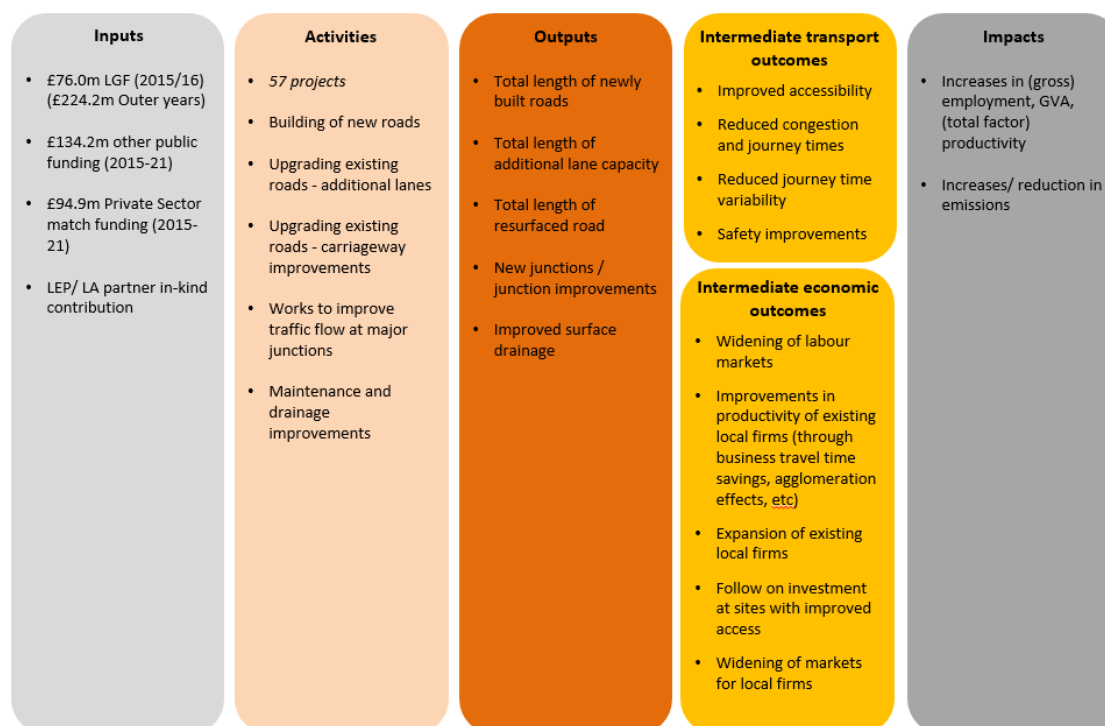
Estimating Gross Impacts: Transport

Road Improvements

Logic chain and monitoring indicators

A logic chain describing the effects of investment in road improvements is summarised below.

Figure 3: Logic model for road improvements

Inputs	Activities	Outputs	Intermediate transport outcomes	Impacts
• £76.0m LGF (2015/16) (£224.2m Outer years) • £134.2m other public funding (2015-21) • £94.9m Private Sector match funding (2015-21) • LEP/ LA partner in-kind contribution	• *57 projects* • Building of new roads • Upgrading existing roads - additional lanes • Upgrading existing roads - carriageway improvements • Works to improve traffic flow at major junctions • Maintenance and drainage improvements	• Total length of newly built roads • Total length of additional lane capacity • Total length of resurfaced road • New junctions / junction improvements • Improved surface drainage	• Improved accessibility • Reduced congestion and journey times • Reduced journey time variability • Safety improvements	• Increases in (gross) employment, GVA, (total factor) productivity • Increases/ reduction in emissions
			Intermediate economic outcomes • Widening of labour markets • Improvements in productivity of existing local firms (through business travel time savings, agglomeration effects, etc) • Expansion of existing local firms • Follow on investment at sites with improved access • Widening of markets for local firms	

Based on this logic model and the suggested monitoring indicators proposed by BEIS, relevant metrics for data collection and analysis are summarised in the table below.

Table 1: Suggested monitoring indicators for investment in road improvements

OUTPUTS		
Measures	**Metrics / Units**	**Potential data sources**
Total length of newly built roads	Km	Monitoring information

Total length of additional lane capacity	Km	Monitoring information
Total length of improved and/or resurfaced road	Km	Monitoring information
New junctions / junction improvements	Number	Monitoring information
Improved surface drainage	Number	Monitoring information

INTERMEDIATE TRANSPORT OUTCOMES

Measures	Metrics / Units	Potential data sources
Improved accessibility: Total number of households able to access site within 20/40 minutes using public transport\walking, car and cycle	Absolute number of people and percentage of households within Local Authority	DfT accessibility statistics / Local Authority data (LTP indicator) https://www.gov.uk/government/statistical-data-sets/acs05-travel-time-destination-and-origin-indicators-to-key-sites-and-services-by-lower-super-output-area-lsoa
Reduced congestion	Traffic count / Average speed on road	DfT traffic counts: http://www.dft.gov.uk/traffic-counts/ DfT congestion and reliability statistics: https://www.gov.uk/government/collections/road-congestion-and-reliability-statistics
Reduced journey times	% of journeys that are 'on time', comparing journey times with historical data for individual sections of road.	DfT traffic counts: http://www.dft.gov.uk/traffic-counts/ DfT congestion and reliability

		statistics
		https://www.gov.uk/government/collections/road-congestion-and-reliability-statistics
Reduced journey time variability	Standard deviation of AM and PM peak hour journey time	DfT congestion and reliability statistics
		https://www.gov.uk/government/collections/road-congestion-and-reliability-statistics
		Trafficmaster data
Safety improvements	Road accident data by mode	DfT Road Safety data (personal injury accidents reported to police: STATS 19 data)
		http://data.gov.uk/dataset/road-accidents-safety-data

INTERMEDIATE ECONOMIC OUTCOMES

Measures	Metrics / Units	Potential data sources
Widening of labour markets	Travel-to-work flows	Annual Population Survey/ primary research
		http://www.ons.gov.uk/ons/rel/regional-trends/area-based-analysis/commuting-patterns-from-the-annual-population-survey--local-authorities--2010-and-2011/index.html
Improvements in productivity of existing local firms (through business travel time savings, agglomeration effects, etc.)	Total factor productivity	Primary research
Expansion of existing local	Employment/ turnover/	BSD/ primary research

firms	investment (£)	http://www.ons.gov.uk/ons/about-ons/business-transparency/freedom-of-information/what-can-i-request/virtual-microdata-laboratory--vml-/index.html
Follow on investment at sites with improved access	Employment/ turnover/ investment/ number of plants (£)	Monitoring information/ primary research
Widening of markets for local firms	Qualitative description / value (£)	Primary research

The monitoring information collected by scheme sponsors for road transport improvements will be useful in terms of determining what the Local Growth Funds have been used for but, being focussed on outputs and results in terms of intermediate transport outcomes for larger projects, will be of limited direct value to the evaluation of the economic development impacts of the projects concerned. A key exception would be any potential data which can be collected on follow-on investment at sites with improved access.

The assessment of intermediate transport outcomes can potentially draw upon the secondary data sources held by DfT. Any modelling based evaluation of the wider economic impacts of transport projects can potentially make use of the Department's accessibility and transport connectivity data which may be more useful than scheme level monitoring data for this purpose. DfT data may in some cases need to be complemented by Local Enterprise Partnerships' (LEPs) own data on intermediate transport outcomes.

Data on intermediate economic outcomes is likely to have more restrictions in terms of the use of secondary data sources, although changes in travel-to-works flows at the local authority level can be measured through the Annual Population Survey. Commuter flows for smaller geographies will be available from the 2011 Census, but this has the obvious limitations of not being able to facilitate a year-on-year analysis.

Longitudinal records of a number of performance metrics (such as employment and turnover[12]) for individual firms and/ or firms located within a defined area will be available through secondary data sources such as the Business Structure Database (BSD) (at least those that are VAT/ PAYE registered). Datalinking should therefore make it possible to estimate any expansion of existing firms. Spatial coverage is available to Middle Super Output Area (MSOA) level with a breakdown by size-band, sector and legal status. Access to the BSD is restricted and requires a formal application/ fee submission.

Datalinking could also be used to measure changes in productivity, although the data required is only available for large firms. As an alternative, turnover per employee (available from the BSD) can be used as a proxy for productivity. Productivity data may also be collected through primary research (i.e. through information on total employment

---

[12] Although this is only available for enterprises and not local units.

costs and profits; and/ or turnover and expenditure on intermediate goods and services to approximate gross value added (GVA)). It is, however, more challenging to gather robust information from surveys than administrative information.

Potential quasi-experimental and other modelling options

Options and associated issues

There is significant variation in the nature of proposed road schemes. Around 32% of funding is directed towards flow improvements at major junctions and roundabouts. Another 28% is designed to improve carriageways with 23% aimed at maintenance and drainage, 13% at road construction (including link roads) and 4% aimed at improving flow management.

Ultimately, the intention of any assessment will be to examine whether these investments have any identifiable effect on the metrics selected to infer impact – in this case local business turnover, employment and productivity or in the numbers of firms operating in the vicinity of the schemes.

A complication in this task, as with all transport evaluations, is the potential issue of reverse causality (otherwise called simultaneity, endogeneity or bi-directionality) whereby the location of investment may be as much influenced by economic factors (positive or negative) as the other way around. This makes it difficult to disentangle impacts that might relate to the investment from other more general area factors.

Assessment: Feasibility

(1) Regression Discontinuity Design (RDD)

There is no general reason why an RDD approach should not be feasible as an ex-post vehicle through which to assess road investment schemes. As long as project applications are approved/scored according to a common process (e.g. benefit-cost ratios (BCRs)), details of both successful and unsuccessful applications are available and there exists a threshold for approval that is quasi-random, then there exists (subject to confirmation of other criteria) a prima-facie basis for RDD.

In practice, the requirements for effective implementation of RDD are quite demanding and there is a potential debate as to whether different RDD models should be applied to different types of road investment schemes. We suspect that it may be difficult to make the case that adequacy requirements will have been met – projects are very varied, apply in very different contexts with no guarantee that all projects will have been assessed in a consistent manner across a set of criteria and ranked and approved or rejected in accordance with such criteria. Likewise, application by scheme may well struggle to produced adequate number of comparators. In this light, we set this option aside.

(2) Instrumental Variables (IV)

There is a history of using IV methods in road or highway impact assessments. An IV approach – defining an instrument which explains selection (location of investment) without also influencing outcomes following the investment – may therefore present a technically feasible approach. On the other hand, traditional applications of IV approaches to road investment are typically at a high level of geography in the context of broad measures of transport infrastructure in which event some vintage infrastructure or transport plan is used as the instrument. More recent research has developed an approach using a truncated or frozen accessibility measure as an IV in the context of accessibility-based assessment which provides a potential alternative and it may be feasible to consider whether DfT accessibility indices might provide a similar function. An IV or IV related approach thereby offers some scope for consideration.

(3) Control group approaches

It is not inconceivable that a control group approach might be considered for assessment of road related investment but the level of detail against which to define controls – junctions, roads, roundabouts, sites with similar traffic levels, functions, capacity, mode mix, local area characteristics, proximity and underlying accessibility – makes it difficult to believe that this would be a successful process in terms of balancing and common support requirements either at an aggregate or a scheme by scheme level.

(4) Control group approaches: Boundary analysis

In contrast to (3), there is the possibility of using a boundary or treatment/intensity approach and to examine performance metrics of businesses both in the immediate vicinity of the investment (most likely to 'treated') and across extending rings surrounding the investment location, having controlled for variation in observables (e.g. firm-level or area-level characteristics). This approach does not require the formal definition of a control group and can be adopted either on a scheme-by-scheme basis or across all schemes.

(5) Dose-response models

Given that there is substantial variation in support levels across projects, and the policy variable can be defined in continuous form (spend), it is feasible to consider whether a dose-response structure would be suited to the analysis of scheme impact.

(6) Accessibility models

A series of recent research papers have developed an approach to the impact of road investment through an accessibility (effective density) measure. It is argued that new roads increase accessibility to employment within a given unit of travel time and that (subject to a number of other adjustments) one can assess whether change in business performance follows change in accessibility. It strikes us that an accessibility approach might prove suitable for consideration.

Assessment: Robustness

The issue of robustness revolves around whether there exists an appropriate and credible identification strategy through which causality can be assured and policy impact defined. In what follows we review robustness only for those options that are deemed feasible and thereby set to one side options (1) and (3).

(2) Instrumental Variables (IV)

Developing appropriate instruments is challenging. Many of the schemes are relatively localised, in particular geographies/local contexts, and we suspect that there will be next to no potential to use a vintage infrastructure plan as the basis of an IV approach.

Potential for constructing a 'frozen' DfT-based accessibility indicator exists but these typically reference the location of service facilities and change in the distribution of such facilities can influence the accessibility measure as much as any change in access due to road conditions or infrastructure. The one exception is in relation to town centres which, of course, remain fixed. Since not all investments relate to town centre accessibility, this may ultimately prove to be of limited value and our understanding is that construction of the indicators is very resource-intensive.

As noted, recent research has developed an accessibility-related instrument specific to the context of road investment which may well prove viable. Discussion is referred to the review of option (6) below, rather than here.

(4) Control group approaches: Boundary analysis

Identifying firms in the vicinity of investments and in surrounding rings is straightforward although some geographic information system (GIS) input may be required to map isochrones onto administrative maps. As the impact datasets of interest are longitudinal in nature, fixed effects panel methods are available and it will be feasible to employ DiD approaches. As such, it is likely that this option will achieve a level 3 SMS score.

(5) Dose-response models

Advances in dose-response methods make it possible to consider this approach. The development of generalised propensity score (GPS) matching means that it is feasible to test for balancing and overlap of covariates, as per PSM. Recent variants of the approach have introduced linear mixed model approaches that are reported to address issues of time-invariant unobservables and bi-directionality between treatment and response. In this case it is possible that the approach would score at level 4 on the SMS scale but development is at an early stage and the transport example of which we are aware operates at a higher level geography and is technically complex[13].

The application of dose-response models to transport evaluation has been primarily in the context of road improvements, it is important to recognise that the approach focusses on interventions designed to expand network capacity rather than the myriad of different

---

[13] See Graham et al (2014).

schemes. It is also important to recognise that this approach is intrinsically different to the detailed accessibility-based studies reviewed elsewhere in this report.

One of the most obvious differences relates to the level of geography employed. In contrast to the accessibility studies where the focus is on relatively low level spatial units (wards), the application is at a relatively high level of geography (U.S. cities). Operating at higher geographies has obvious potential benefits in terms of data management but construction of (aggregate) datasets connected to lower-level interventions is not necessarily straightforward. In this case covariates include traffic volume, traffic and network composition (freeway/arterial), network scale, mode share, congestion, economic structure (% jobs in manufacturing), employment and population level/growth, income and state fuel prices. Performance variables are defined as annual proportional change in traffic volume, network performance (annual hours of delay) and productivity (average wages).

The approach also extends existing Generalised Propensity Score (GPS) methods into a longitudinal framework, not only accommodating (observed) time-varying and (unobserved) time-invariant characteristics/features but bi-directionality. This level of robustness is achieved by using advanced statistical methods and our understanding is that the routines required to implement these new techniques were custom-developed by the authors. This makes replication in other contexts, by other parties, difficult.

While there is no intrinsic impediment to adoption of a similar approach for the evaluation of transport interventions, there are clear challenges in so doing. In the first instance, the varying nature of LEPs and their differing transport contexts means that there is much less homogeneity than in the case of a city-level study. It might be an option to construct a pan-LEP city-level dataset but this would also require generating specific data series and it is not clear whether the inevitably smaller sample size would in any way restrict application of the techniques. Secondly, as indicated, the routines required to implement the techniques are (at this point in time) proprietary in nature. Finally, applied in the context of a twenty-five year time-span, it is not clear whether the approach is robust to implementation in a shorter time-frame more appropriate to evaluation of interventions.

(6) Accessibility models

Research demonstrates that accessibility-based approaches can provide a mechanism for evaluating investment in road capacity[14]. Whether a similar approach is capable of performing the same task here is unproven but we are strongly of the view that it will prove fit for purpose in that there exists a methodology that addresses many of the problems that diminish robustness of transport evaluations and that can operate at a low level geography.

That said, it is also clear that some further development is required before such an approach could be employed. In particular, while some schemes relate to additional capacity, much of the investment relates to surface accessibility – alleviating bottlenecks to reduce time taken to travel between points. We can see no technical reason why the

---

[14] See Gibbons et al (2012) and Sanchis-Guarner (2013).

accessibility-based approach on capacity should not be extended to incorporate surface access considerations and used in the same way to examine potential impacts[15].

Operating at individual LEP level (i.e. only those LEPs that receive investment rather than across the country as a whole), this task would require definition of an origin-destination matrix (ODM)[16], application of an optimal routing algorithm to determine travel times on an annual basis prior to and post investment, and definition of an accessibility measure within a set time/distance limit of each origin/destination, also calculated on an annual basis.

Undertaken in the same manner as the research studies, the approach would also require development of a truncated/frozen accessibility measures as an instrument – this simply means 'freezing out' the impact of employment change from the calculation of the accessibility measure prior to the investment[17].

Differencing and applying fixed effects methods, with impact metrics assessed through a series of bands around the investment locations, should enable a minimum level 3 on the SMS scale.

The approach, if viable, would also be capable of accommodating multiple and overlapping investments as they would be referenced in the same accessibility framework and reflected in the same accessibility measure. Moreover, if other transport investments (rail, light rail) could be assessed similarly, there may be scope to consider a combined multi-modal framework.

Assessment: Overview & Implementation

The feasibility and robustness commentary indicates that priority should initially be given to option (6) with option (4) as a fall-back. As indicated, this represents a relatively novel modus operandi and in terms of implementation, it will be necessary for evaluators to:

- Liaise with sponsors in order to:
  - o understand the nature and nuances of the interventions delivered;
  - o define the specific (geocoded) locations of investments;
- Construct an accessibility framework:
  - o identify a convenient spatial unit within which to operate (e.g. wards);
  - o define an origin-destination matrix (ward to ward) for each LEP subject to investment, including across LEP boundaries if investments are close to the boundary edge[18];
  - o apply a routing algorithm to the ODM for a base year and period prior to the investment and annually thereafter to provide an annual record of journey

---

[15] The extension would represent an attempt to incorporate 'flow features' such as congestion into the framework. If the investments primarily relate to the major road network, it may be the case that the DfT AADF (Annual average daily flow) statistics can provide a basis for this adjustment.

[16] i.e. a series of sub-geographies covering the LEP that can be used as the basis of journey time calculations. It might be simplest to follow the research and use wards.

[17] One departure from the studies might relate to their choice to eliminate areas close to any investment site from consideration of impact. There is a debate here about whether the reasoning for this action in the context of major road schemes should be applicable to much more modest localised investments.

[18] There may be an issue of road investments in neighbouring LEPs falling within another LEP matrix if working across boundaries but it is not considered here.

times from any origin to a given destination and taking into account reliability[19];
- o construct an accessibility measure for each spatial unit (ward) included in the analysis[20];
- o construct a variant of the latter with employment fixed at a point prior to investment to act as an instrument;
- Construct an impact geography:
  - o map investment locations to OD accessibility profile;
  - o define rings/bands of different distances from the site of each investment;
- Cross-reference with the BSD/ARD (Annual Respondents Database) and extract:
  - o details of performance (turnover, employment, productivity, firm numbers) of each business pre/post intervention within each of the control rings;
- Specify estimation model and assess impact:
  - o define the nature of the reduced form model that is to be estimated across rings;
  - o taking account of any controls/variables (firm characteristics, area and local economy characteristics) believed likely to impact on performance profiles;
  - o using fixed effects/differencing and IV techniques.

Potential alternative and supplementary approaches

Two aspects need to be considered: assessing the economic development contribution and including the transport efficiency effects (e.g. travel time savings).

Economic development contribution

Potential means of assessing the economic development contribution of projects should recognise:

a) In some cases it will be possible to secure direct evidence that highway schemes have played a crucial role in allowing the development and employment change observed through the monitoring process to go forward:

- The development involved could not physically have taken place without the improvements involved;
- It would otherwise likely have been blocked through the planning process by objections by the local highway authority or by the Highways Agency through Article 15 Directions.

It is suggested that the LEPs should be asked to bring forward and assess (any) relevant evidence of this sort as part of their local evaluation plans.

---

[19] There are differing levels of complexity here regarding measures of reliability and timeframes to be employed in any calculation.
[20] Ultimately, this may produce a figure which is similar in some ways to the DfT accessibility measure but referencing wards rather than employment/service centres.

b) In at least some cases the wider economic impacts of projects will have been subject to an economic benefits assessment (potentially considering aspects such as imperfect competition, labour supply and agglomeration impacts) and/or an Economic Impact Report (EIR) will have been prepared in accordance with DfT WebTAG guidance.

In either case it is suggested that the validity of these assessments should be reconsidered as part of the LEP local evaluation plans. In the case of the wider economic benefits the assessment will need to involve rerunning the original modelling based upon actual post-completion traffic data. In the case of the EIRs a more eclectic approach will be required, reflecting the more diverse focus and methods of such assessments. It may be appropriate to limit this to some case studies of larger projects.

It is unclear at this stage whether there are any Highways Agency projects within the first round of Deals which will be subject to post-completion (POPE) evaluations under the Agency's normal criteria. If there are, there will clearly be a need for co-ordination.

It should be noted that the impacts identified under approaches a) and b) will not generally be additive to those measured through the spatial modelling. Rather they represent a complementary source of evidence to be 'triangulated' with the spatial modelling evidence in arriving at overall impact assessments. However, it needs to be acknowledged that the agglomeration and other wider economic impacts may partly arise outside the impact area though, in practice, this will probably need to be ignored as a 'second order' problem.

Transport efficiency aspects

Transport efficiency aspects, including travel time and accident cost savings, will need to be taken into account in any overall economic evaluation of the Deals. They will need to be assessed via either one or a combination of:

- Utilisation of the available ex-ante modelling evidence;
- Where practicable and proportionate for the larger projects, reassessment of this evidence based upon post-completion travel time and accident data.

There will clearly be some double counting between the productivity benefits to existing local businesses and the work based travel time savings as estimated through the transport efficiency analysis. It is suggested that – assuming clear evidence of local productivity impacts is identified and built into the appraisal – it may be appropriate to discount the proportion of the work based travel time savings from the transport efficiency assessment which are associated with journeys by employees of the businesses concerned, at least for the period over which the productivity impacts are assumed to persist.

Summary and conclusions – Road Improvements

The proposed approach here builds upon recent academic studies. It involves the construction of pre and post-investment matrices of journey times by road to identify potentially significant travel time savings between relevant origin/destination pairs. These would then feed as an instrument, along with relevant control variables, into a fixed effects modelling framework designed to assess the extent to investments 'explain' variations in small area economic performance between locations at varying distances from the schemes involved as a basis for assessing their impacts.

There needs to be a recognition that the nature and limited scale of many of the investments involved and the potential issues in incorporating effects in alleviating congestion create a particularly challenging context for the application of this method. For this reason and to provide a richer evidence base with which the modelling results can be triangulated, it is suggested that the LEPs should be required to collect supporting evidence on local impacts as part of their own evaluation activity. There will also be a need to assemble evidence on transport efficiency aspects to support the overall economic evaluation, at least for the more major schemes.

Urban Sustainable Transport

Scope of interventions

Growth Deal investment of £689m in 78 urban sustainable projects is planned between 2015/16 and 2020/21. These projects comprise a mixture of single and multi-modal improvements to improve public transport in urban areas. The LGF funded projects are primarily concerned with modal shift, and generally attempt to create viable transport alternatives with significant investment in cycling/pedestrian paths, rail station improvements, bus lane management, etc.

A large proportion of these projects, particularly the broad multi-modal schemes, focus generally on improving urban sustainable transport. However, some projects specifically targets commuters (i.e. they attempt improve sustainable transport links between major residential areas and places of employment).

There are also a number of projects that attempt to use sustainable transport to improve the attractiveness of areas in order to boost economic development and growth through increased visitor numbers.

Logic chain and monitoring indicators

A logic chain describing the effects of investment in urban sustainable transport is summarised below.

Figure 4: Logic model for urban sustainable transport

Inputs	Activities	Outputs	Intermediate transport outcomes	Impacts
• £159.5m LGF (2015/16) (£529.5m outer years) • £333.9m other public funding (2015/16-2020/21) • £188.0m Private Sector match funding (2015/16-2020/21) • LEP/ LA partner in-kind contribution	• *78 projects* • Investment in multi-modal transport hubs • Bus rapid transit (BRT) schemes • Metro extension works • New bus station works • Investment in Intelligent Transport Systems	• Type of service improvement (e.g. provision of traffic management information) • Type of infrastructure (e.g. transport interchanges/ hubs, new metro links/ bus stations/ railway links)	• Improved accessibility • Increased usage of non-car modes/ change in mode split • Highway decongestion/ possible reduced journey times/ reliability effects • Safety improvements **Intermediate economic outcomes** • Widening of labour markets • Improvements in productivity of existing local firms (through business travel time savings, agglom. effects, etc) • Expansion of existing local firms • Follow on investment at/ near transport nodes • Possible effects on commercial rentals	• Increases in (gross) employment, GVA, (total factor) productivity • Reduction in emissions

Based on this logic model and the suggested monitoring indicators proposed by BEIS, relevant metrics for data collection and analysis are summarised in the table below.

Table 2: Suggested monitoring indicators for investment in urban sustainable transport

OUTPUTS		
Measures	**Metrics / Units**	**Potential data sources**
Type of service improvement	Description of development	Monitoring information
Type of infrastructure	Description of development	Monitoring information
INTERMEDIATE TRANSPORT OUTCOMES		
Measures	**Metrics / Units**	**Potential data sources**
Improved accessibility: Total number of households able to access site within 20/40 minutes using public transport\walking, car and cycle	Absolute number of people and percentage of households within Local Authority	DfT accessibility statistics / Local Authority data (LTP indicator) https://www.gov.uk/government/statistical-data-sets/acs05-travel-time-destination-and-origin-indicators-to-key-sites-

		and-services-by-lower-super-output-area-lsoa
Increase usage of non-car modes/change in mode split	Traffic count by mode	DfT traffic counts: http://www.dft.gov.uk/traffic-counts/
Reduced congestion	Traffic count / Average speed on road	DfT traffic counts: http://www.dft.gov.uk/traffic-counts/ DfT congestion and reliability statistics: https://www.gov.uk/government/collections/road-congestion-and-reliability-statistics
Reduced journey times	% of journeys that are 'on time', comparing journey times with historical data for individual sections of road.	DfT traffic counts: http://www.dft.gov.uk/traffic-counts/ DfT congestion and reliability statistics https://www.gov.uk/government/collections/road-congestion-and-reliability-statistics
Reduced journey time variability	Standard deviation of AM and PM peak hour journey time	DfT congestion and reliability statistics https://www.gov.uk/government/collections/road-congestion-and-reliability-statistics Trafficmaster data

| Safety improvements | Road accident data by mode | DfT Road Safety data (personal injury accidents reported to police: STATS 19 data)

http://data.gov.uk/dataset/road-accidents-safety-data |

INTERMEDIATE ECONOMIC OUTCOMES

Measures	Metrics / Units	Potential data sources
Widening of labour markets	Travel-to-work flows	Annual Population Survey/ primary research http://www.ons.gov.uk/ons/rel/regional-trends/area-based-analysis/commuting-patterns-from-the-annual-population-survey--local-authorities--2010-and-2011/index.html
Improvements in productivity of existing local firms (through business travel time savings, agglomeration effects, etc.)	Total factor productivity	Primary research
Expansion of existing local firms	Employment/ turnover/ investment (£)	BSD http://www.ons.gov.uk/ons/about-ons/business-transparency/freedom-of-information/what-can-i-request/virtual-microdata-laboratory--vml-/index.html
Follow on investment at sites with improved access	Employment/ turnover/ investment/ number of plants (£, by source)	Monitoring information/ primary research
Follow on investment at/near transport nodes	Employment/ turnover/ investment/ number of plants (£, by source)	Monitoring information/ primary research
Effects on commercial rents	The market rate for leasing commercial floorspace at the "impact" site (£/sq m per	Monitoring information/ primary research

	month, by class)	

As well as collecting data on the project outputs, there is also an expectation on scheme sponsors to gather information on any follow on investment that can be linked to the transport improvements. The scheme sponsors are also expected to collect information on commercial rental values at/ near the transport nodes.

However, in reality this may represent a difficult proposition and the evaluation of the projects may therefore need to be complemented by data from primary research.

Secondary data sources published by DfT will provide a strong basis for estimating intermediate transport outcomes, including accessibility, congestion, journey times and safety improvements. DfT data may in some cases need to be complemented by LEPs' own data on intermediate transport outcomes.

The collation of data in respect of intermediate economic outcomes will be similar to that of the road improvement projects described above.

Potential quasi-experimental and other modelling options

Options and associated issues

Urban sustainable transport is the largest of the three transport intervention types with spend over double that of roads and six times greater than for non-urban rail. Projects are primarily targeted at improvements to existing lines/facilities in order to encourage modal shift or transport alternatives though a number also focus on improving commuter links and using sustainable links to promote growth/regeneration.

The intention of any assessment will be to examine whether these investments have any identifiable effect on the metrics selected to infer impact – in this case local business turnover, employment and productivity or in the numbers of firms operating in the vicinity of the schemes. It might also be of interest to examine house price changes.

In common with all transport evaluations, there is the issue of reverse causality whereby the location of investment may be as much influenced by economic factors (positive or negative) as the other way around.

Assessment: Feasibility

(1) Regression Discontinuity Design (RDD)

As with road interventions, there is no reason why an RDD approach should not be feasible as an ex-post mechanism for evaluation of sustainable transport schemes. Once again, however, we suspect that it may be difficult to make the case that adequacy requirements will have been met. In this light, we set this option aside.

(2) Instrumental Variables (IV)

IV methods are less prevalent in light-rail/sustainable transport than in the case of road or rail. There is no reason, other than in the difficulty of identifying an appropriate instrument, why such an approach should not be technically feasible. As noted, applications in other transport contexts tend to be at a high level of geography and in the context of broad measures of transport infrastructure in which event some vintage infrastructure or transport plan is used as the instrument. The development of accessibility-based approaches provides an opportunity to consider accessibility-based IVs.

(3) Control group approaches

There is great variety in the nature of supported projects within this intervention category. Ultimately, however, the larger investments will be intended to improve/extend networks (lines/routes) or nodes (stations/hubs). As such, it may be feasible to adopt a standard control group approach provided that testing for balancing and common support across observables proves acceptable:

station/hub construction/upgrades:

- using stations on the same line but not subject to investment, as a control, drawing boundaries around the stations and using these to assess relative impact;
- using hubs within broadly comparable contexts as a control, drawing boundaries and using these to assess relative impact;

network improvements:

- adopting a stretch of network on the same route, not subject to investment, as a control, drawing boundaries around both links (and/or possibly associated stations) and using these to assess relative impact;
- identifying network links that are comparable in terms of passenger load, frequency, time, proximity and a range of other observable characteristics that could serve as controls.

(4) Control group approaches: Boundary analysis

Alongside (3), there is the possibility of using a boundary or treatment/intensity approach and to examine performance metrics of businesses both in the immediate vicinity of the investment and across extending rings/bands surrounding the investment location, having controlled for variation in observables (firm, area, economy related). This approach does not require the formal definition of a control group but uses differences in 'exposure' to the investment as the basis of assessing impact.

(5) Dose-response models

Given that there is substantial variation in support levels across projects, and the policy variable can be defined in continuous form (spend), it is feasible to consider whether a dose-response structure would be suited to the analysis of scheme impact.

(6) Accessibility models

There is the option of developing accessibility related models using a specifically constructed access measure and assessing whether change in business performance follows change in measured accessibility. Again, it strikes us that an accessibility approach is a strong candidate for consideration. In the case of sustainable transport there are two potential accessibility variants - the first is a 'surface access' measure while the second builds upon the DfT accessibility to town centres measure.

Assessment: Robustness

The issue of robustness revolves around whether there exists an appropriate and credible identification strategy through which causality can be assured and policy impact defined. In what follows we review robustness only for those options that are deemed feasible and thereby set to one side option (1).

(2) Instrumental Variables (IV)

Some of the investments relate to (current/previous) large-scale infrastructure and it may be the case that there do exist plans of sufficient vintage to act as an IV. Whether this is the case would require investigation in a scheme by scheme basis but it is doubtful whether this will be the case across all schemes. As noted, recent research has developed an accessibility-related instrument specific to the context of road investment which may well prove replicable in the case of sustainable transport. Discussion of this and consideration of a frozen DfT accessibility indicator is referred to the review of option (6) below, rather than here.

(3) Control group approaches

The issue of relevance in all control group approaches is the quality of the comparator(s) and the existence or otherwise of statistical equivalence on observables. Identifying networks or nodes that are equivalent in terms of a range of characteristics is always going to be challenging though it is, ultimately, an empirical matter which bears some degree of risk. Adopting the approach which takes existing, but non-targeted networks/nodes as comparators may be more convenient from both methodological and empirical perspectives.

Access to performance metrics for either approach is relatedly straightforward with an ability to establish a longitudinal dataset both pre/post investment and apply of panel/DiD techniques. As outlined, this approach does not address the issue of endogeneity and may score at level 3 on the SMS scale but risks level 2 in the event that suitable matching characteristics cannot be demonstrated.

(4) Control group approaches: Boundary analysis

Identifying firms in the vicinity of investments and in surrounding rings is straightforward although some GIS input may be required to map isochrones onto administrative maps. As the impact datasets of interest are longitudinal in nature, fixed effects panel methods are available and it will be feasible to employ DiD approaches. As such, it is likely that this option will achieve a level 3 SMS score.

(5) Dose-response models

Advances in dose-response methods again make it possible to consider this approach. The development of generalised propensity score (GPS) matching means that it is feasible to test for balancing and overlap of covariates, as per PSM. As noted, recent variants of the approach have introduced linear mixed model approaches that are reported to address issues of time-invariant unobservables and bi-directionality between treatment and response. In this case it is possible that the approach would score at level 4 on the SMS scale but, as indicated, development is at an early stage, the transport example of which we are aware operates in the context of roads at a higher level geography and is technically complex.

(6) Accessibility models

In common with the case of roads, we are of the view that there is a case for exploring accessibility models as a vehicle for impact assessment of rail investment.at a low level geography.

(6a) Surface access model

Some development would be required before such an approach could be employed but we can see no technical impediment to construction of a surface access measure which might be used to examine potential impacts.

Operating once more at individual LEP level (i.e. only those LEPs that receive investment rather than across the country as a whole), this task would require definition of an origin-destination matrix (ODM)[21], definition of travel times on an annual basis prior to and post investment, and construction of an appropriate surface accessibility measure within a set time/distance limit of each origin/destination also calculated on an annual basis.

Undertaken in the same manner as the research studies, the approach would also require development of a truncated/frozen accessibility measures as an instrument – 'freezing out' the impact of employment change from the calculation of the accessibility measure prior to the investment[22].

Development of a truncated/frozen accessibility measure as an instrument, differencing and applying fixed effects methods, with impact metrics assessed through a series of bands around the investment locations/routes, should enable a minimum level 3 on the SMS scale.

The approach, if viable, would also be capable of accommodating multiple and overlapping investments as they would be referenced in the same accessibility framework and reflected in the same accessibility measure. In addition, if the approach were to operate

---

[21] i.e. a series of sub-geographies covering the LEP that can be used as the basis of journey time calculations. It might be simplest to follow the research and use wards.
[22] As with roads, one departure from the studies might relate to their choice to eliminate areas close to any investment site from consideration of impact.

using the same ODM units then there may be scope to consider a combined multi-modal framework across the same spatial area (e.g. LEPS).

(6b) Town Centre access model

DfT accessibility indicators examine accessibility to certain service facilities (doctors, petrol stations and so on) including town centres, by LSOA. At first sight, the indicators appear to offer potential for development of an accessibility measure suited to sustainable transport schemes intended to improve access to town centres.

This is especially the case as town centre locations are fixed, unlike other service centres for which the statistics are calculated. This indicator should thereby provide a consistent indication of change in accessibility due to transport improvements in terms of public transport, car travel and cycling.

Potential to use a 'frozen' DfT-based accessibility indicator, specific to light-rail and related investments, is complicated by the fact that the latter use a generic public transport definition. Whether a light-rail component can be isolated and extracted requires further investigation but it may be that constructing an alternative would be time-consuming and resource intensive.

Assessment: Overview & Implementation

The feasibility and robustness commentary indicates that priority should be initially be given to option (6a) with option (4) as a fall-back. As indicated, this represents a relatively novel modus operandi and in terms of implementation, it will be necessary for evaluators to:

- Liaise with sponsors in order to:
  o understand the nature and nuances of the interventions delivered;
  o define the specific (geocoded) locations of investments;
- Construct an accessibility framework:
  o identify a convenient spatial unit within which to operate (e.g. wards);
  o define an origin-destination matrix for each LEP, including across LEP boundaries if investments are close to the boundary edge[23];
  o apply a routing algorithm to the ODM for a base year and period prior to the investment and annually thereafter to provide an annual record of journey times from any origin to a given destination and taking into account reliability[24];
  o define travel times for a base year and period prior to the investment and annually thereafter;
  o construct an accessibility measure for each spatial unit (ward) included in the analysis[25];

---

[23] There may be an issue of sustainable transport investments in neighbouring LEPs falling within another LEP matrix if working across boundaries but it is not considered here.
[24] There are differing levels of complexity here regarding measures of reliability and timeframes to be employed in any calculation.
[25] Ultimately, this may produce a DfT accessibility type measure but referencing wards rather than employment/service centres.

- o construct a variant of the latter with employment fixed at a point prior to investment to act as an instrument;
- Construct an impact geography:
  - o map investment locations to OD accessibility profile;
  - o define rings/bands of different distances from the site of each investment;
- Cross-reference with the BSD/ARD and extract:
  - o details of performance (turnover, employment, productivity, firm numbers) of each business pre/post intervention within each of the control rings;
- Specify estimation model and assess impact:
  - o define the nature of the reduced form model that is to be estimated across rings;
  - o taking account of any controls/variables believed likely to impact on performance profiles;
  - o using fixed effects/differencing and IV techniques.

Potential alternative and supplementary approaches

The major concerns here are whether the available secondary transport and Valuation Office Agency (VOA) data will be available at a sufficiently fine-grained level to allow impacts on mode of travel and rental values to be identified through the modelling process suggested above. The presumption is that use may need to be made instead of locally assembled monitoring data to:

- Identify effects on mode of travel along corridors and in centres benefitting from the schemes being funded through the Deals;
- Identify associated effects on journey times/congestion and accident rates; and
- Identify localised effects on land values, rentals and physical development schemes in the corridors/centres concerned.

As suggested previously, LEPs could be asked to propose suitable comparison corridors/centres and to assemble relevant monitoring data for these as part of their evaluation plans. If identifying suitable comparators within the relevant LEP areas proves problematic, as well may be the case, a strategy may be needed to secure comparators from elsewhere as part of the national evaluation.

As proposed in relation to road schemes, evidence will be needed on the transport efficiency benefits of the investments involved. It seems unlikely that the necessary evidence will be available for all schemes and creative approaches to 'gap filling' may be required – for example, through grossing up evidence from comparable schemes based upon expenditure data.

Potential double counting of local productivity effects and transport efficiency benefits seems unlikely to be a major issue in this case.

Summary and conclusions – Urban Sustainable Transport

The proposed approach broadly follows that proposed in relation to roads, involving the construction of matrices of pre and post-investment journey times by public transport between relevant origin/destination (o/d) pairs and their development into an instrumental variable to be used in a fixed effects model to assess the extent to which the investments explain variations in economic performance between locations at varying distance from the projects concerned.

This represents a relatively novel application of the approach and there are questions about whether/how far aspects such as changes in service frequency/quality and improvements in opportunities for walking and cycling can be taken into account. However, there is a potential variation based upon the use of the DfT Access to Town Centres Indices as an alternative instrumental variable if the construction of the o/d matrices proves problematic.

Reflecting both inevitable uncertainties about the extent to which the modelling will be able to identify what may be limited and complex effects on local economic activity and doubts about the likely usefulness of secondary data on important metrics such as rentals and local development activity, it is suggested that LEPs should be required to collect monitoring data on these metrics for both the treatment areas/corridors and potential comparators as part of the local evaluation effort. Again evidence on transport efficiency benefits will be needed to feed into the economic evaluation.

Rail Transport

Scope of interventions

Growth Deal investment of £55m in nine rail improvement projects is planned between 2015/16 and 2020/21. These projects comprise a mixture of station and rail track improvements, which are designed to either improve rail capacity and frequency of service or enhance speed and reliability of the service. For example:

- Three projects provide general station upgrades (e.g. station enhancements and better track and signalling) aimed at improving access, capacity and service frequency.
- Three projects focus on rail route improvements. These include interventions that improve speed and efficiency, such as track electrification, as well as a project that increases passenger capacity with a more regular service.
- Two projects concern the extension and reopening of rail routes, thus improving capacity.
- One project has been funded to undertake a study into improvements and re-modelling of a rail station.

Logic chain and monitoring indicators

A logic chain describing the effects of investment in rail transport improvements is summarised below.

Figure 5: Logic model for rail transport improvements
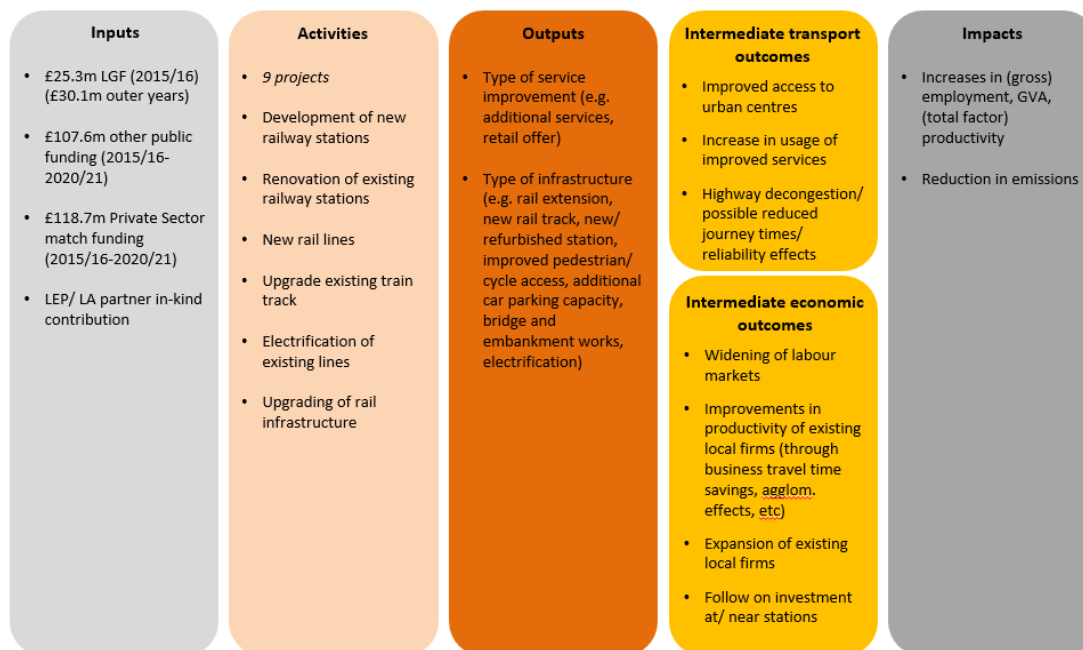
Based on this logic model and the suggested monitoring indicators proposed by BEIS, relevant metrics for data collection and analysis are summarised in the table below.

Table 3: Suggested monitoring indicators for investment in non-urban rail improvements

OUTPUTS

Measures	Metrics / Units	Potential data sources
Type of service improvement	Description of development	Monitoring information
Type of infrastructure	Description of development	Monitoring information

INTERMEDIATE TRANSPORT OUTCOMES

Measures	Metrics / Units	Potential data sources
Improved access to urban centres	Households with access to specific urban centres within 20/40 minutes using public transport	DfT accessibility statistics / Local Authority data (LTP indicator) https://www.gov.uk/governme

		nt/statistical-data-sets/acs05-travel-time-destination-and-origin-indicators-to-key-sites-and-services-by-lower-super-output-area-lsoa
Increased in usage of improved services	Annual average daily passenger boardings; AM, inter- and PM peak hour passenger boardings	ORR/ primary research http://orr.gov.uk/statistics/data-portal
Highway decongestion / possible reduced journey times / reliability effects	AM and PM peak proportion of trips for different travel modes	DfT congestion and reliability statistics https://www.gov.uk/government/collections/road-congestion-and-reliability-statistics

INTERMEDIATE ECONOMIC OUTCOMES

Measures	Metrics / Units	Potential data sources
Widening of labour markets	Travel-to-work flows	Annual Population Survey/ primary research http://www.ons.gov.uk/ons/rel/regional-trends/area-based-analysis/commuting-patterns-from-the-annual-population-survey--local-authorities--2010-and-2011/index.html
Improvements in productivity of existing local firms (through business travel time savings, agglomeration effects, etc.)	Total factor productivity	Primary research
Expansion of existing local firms	Employment/ turnover/ investment (£)	BSD http://www.ons.gov.uk/ons/about-ons/business-transparency/freedom-of-information/what-can-i-request/virtual-microdata-laboratory--vml-/index.html

Follow on investment at sites with improved access	Employment/ turnover/ investment/ number of plants (£, by source)	Monitoring information/ primary research
Follow on investment at/ ear transport nodes	Employment/ turnover/ investment/ number of plants (£, by source)	Monitoring information/ primary research

The evaluation on rail transport projects will draw on similar data sources (secondary and primary) to those detailed above. However, it will also be able to draw on a range of additional data sources. For example, rail passenger numbers (and associated data on overcrowding) can be derived from passenger counts undertaken by train operators at points along routes either manually (on-board by guards, platform) or electronically (load weighting, infra-red). It is, however, clear that there is significant scope for sampling variation and measurement error to occur in generation of passenger number statistics.

The Office for Rail Regulation (ORR) also produces statistics on passenger rail use with data obtained from the Association of Train Operators (ATOC), the rail industry ticketing and revenue database (LENNON) and Train Operating Companies (TOCs). Generation of activity is a complex calculation involving mathematical modelling of ticket sales and passenger profiles are produced for three area sectors – London and South East, long distance and regional. There are, however, a number of limitations that exist in generating the dataset with Steer Davies Gleave commenting (in preparing the datasets and with reference to LENNON) that 'the data is being pushed significantly beyond what it was originally designed for'. In addition, while the dataset provides a relatively local perspective on station use, methodological changes constrain year-on-year comparisons.

On account of these limitations, it may ultimately be more productive to seek the assistance of relevant operators in generating relevant and timely intelligence.

Potential quasi-experimental and other modelling options

Options and associated issues

Non-urban rail is the smallest of the three transport intervention categories with spend two thirds smaller than for roads and 80% lower than urban sustainable transport. Both station upgrades to improve access/frequency and route improvements each account for just under 50% of funding. The remainder is primarily allocated for purposes of extending and reopening of routes.

As with the other transport schemes, the intention of any assessment will be to examine whether these investments have any identifiable effect on the metrics selected to infer impact – in this case local business turnover, employment and productivity or in the numbers of firms operating in the vicinity of the schemes. It might also be of interest to examine house price changes.

Similarly, there is the issue of reverse causality whereby the location of investment may be as much influenced by economic factors (positive or negative) as the other way around.

The development of accessibility-based approaches provides an opportunity to consider accessibility-based IVs.

Assessment: Feasibility

(1) Regression Discontinuity Design (RDD)

There is no reason why an RDD approach should not be feasible as an ex-post mechanism for evaluation of rail investment schemes. In common with road and sustainable transport, however, we suspect that it may be difficult to make the case that adequacy requirements will have been met and the generally low numbers of schemes will limit application. We therefore set this option aside.

(2) Instrumental Variables (IV)

There does exist a strand of research and studies that use IV methods in rail impact assessments. An IV approach may therefore present a technically feasible approach. As elsewhere, however, successful applications are typically at a high level of geography and employ some vintage infrastructure or transport plan as an instrument. It may be possible to relate some schemes to vintage plans and the development of accessibility-based approaches provides an opportunity to consider accessibility-based IVs.

(3) Control group approaches

Rail projects are essentially one of two kinds - station upgrades to improve access/frequency and route improvements. Relatively limited in number, it may be feasible to adopt a standard control group approach provided that testing for balancing and common support across observables proves acceptable:

station construction/upgrades:

- using stations on the same line but not subject to investment, as a control, drawing boundaries around the stations and using these to assess relative impact;
- identifying a station that is comparable in terms of passenger load, frequency, time, proximity and a range of other observable characteristics that could serve as a control.

routeline improvements:

- adopting a stretch of line on the same route, not subject to investment, as a control, drawing boundaries around both links (and possibly associated stations) and using these to assess relative impact;
- identifying a rail link that is comparable in terms of passenger load, frequency, time, proximity and a range of other observable characteristics that could serve as a control.

(4) Control group approaches: Boundary analysis

There is the possibility of using a boundary or treatment/intensity approach and to examine performance metrics of businesses both in the immediate vicinity of the investment and

across extending rings surrounding the investment location, having controlled for variation in observables. As noted previously, this does not require the formal definition of a control group but uses differences in 'exposure' to the investment as the basis of assessing impact.

(5) Dose-response models

Given that there is substantial variation in support levels across projects, and the policy variable can be defined in continuous form (spend), it is feasible to consider whether a dose-response structure would be suited to the analysis of scheme impact.

(6) Accessibility models

There is the option of adopting a surface accessibility measure discussed preciously and as demonstrated in the case of roads by recent research studies. This would effectively require the construction of rail accessibility measure and assessing whether change in business performance follows change in measured accessibility. Again, it strikes us that an accessibility approach is a strong contender for consideration.

Assessment: Robustness

The issue of robustness revolves around whether there exists an appropriate and credible identification strategy through which causality can be assured and policy impact defined. In what follows we review robustness only for those options that are deemed feasible and thereby set to one side options (1).

(2) Instrumental Variables (IV)

Some of the investments relate to (current/previous) large-scale infrastructure and it may be the case that there do exist plans of sufficient vintage to act as an IV. Whether this is the case would require investigation in a scheme by scheme basis but it is doubtful whether this will be the case across all schemes. As noted, recent research has developed an accessibility-related instrument specific to the context of road investment which may well prove replicable in the case of sustainable transport. Discussion of this and consideration of a frozen DfT accessibility indicator is referred to the review of option (6) below.

(3) Control group approaches

Identifying facilities (such as stations, rail links) that are equivalent in terms of a range of characteristics is always going to be challenging though it is, ultimately, an empirical matter which bears some degree of risk. Adopting the approach which takes existing, but non-targeted stations/links as comparators may be more convenient from both methodological and empirical perspectives.

Access to performance metrics for either approach is relatedly straightforward with an ability to establish a longitudinal dataset both pre/post investment and application of panel/DiD techniques. As outlined, this approach does not address the issue of endogeneity and may score at level 3 on the SMS scale but risks level 2 in the event that suitable matching characteristics cannot be demonstrated.

(4) Control group approaches: Boundary analysis

Identifying firms in the vicinity of investments and in surrounding rings is straightforward although some GIS input may be required to map isochrones onto administrative maps. As the impact datasets of interest are longitudinal in nature, fixed effects panel methods are available and it will be feasible to employ DiD approaches. As such, it is likely that this option will achieve a level 3 SMS score.

(5) Dose-response models

Advances in dose-response methods again make it possible to consider this approach. The development of generalised propensity score (GPS) matching means that it is feasible to test for balancing and overlap of covariates, as per PSM. We have already noted that recent variants of the approach have introduced linear mixed model approaches that are reported to address issues of time-invariant unobservables and bi-directionality between treatment and response. In this case it is possible that the approach would score at level 4 on the SMS scale but development is at an early stage, the transport example of which we are aware operates in the context of roads and at a higher level geography and is technically demanding.

(6) Accessibility models

An accessibility based approach is, in our view, offers the most robust modus-operandi in the context of rail, though this is balanced by a greater complexity which would require some work developing a surface-access measure before such an approach could be employed.

Operating at individual LEP level, this task would require definition of an origin-destination matrix (ODM)[26], definition of travel times on an annual basis prior to and post investment, and construction of an appropriate surface accessibility measure within a set time/distance limit of each origin/destination also calculated on an annual basis.

Undertaken in the same manner as the road research studies, the approach would also require development of a truncated/frozen accessibility measures as an instrument – 'freezing out' the impact of employment change from the calculation of the accessibility measure prior to the investment[27]. Differencing and applying fixed effects methods, with impact metrics assessed through a series of bands around the investment locations/routes, should enable a minimum level 3 on the SMS scale.

The approach, if viable, would also be capable of accommodating multiple and overlapping investments as they would be referenced in the same accessibility framework and reflected in the same accessibility measure. In addition, if the approach were to operate using the same ODM units then there may be scope to consider a combined multi-modal framework across the same spatial area (e.g. LEPS).

---

[26] i.e. a series of sub-geographies covering the LEP that can be used as the basis of journey time calculations. It might be simplest to follow the research and use wards.
[27] As with roads, one departure from the studies might relate to their choice to eliminate areas close to any investment site from consideration of impact.

Assessment: Overview & Implementation

The feasibility and robustness commentary indicates that priority should be initially be given to option (6) with option (4) as a fallback. Once again, this represents a relatively novel modus operandi and in terms of implementation, it will be necessary for evaluators to:

- Liaise with sponsors in order to:
  - understand the nature and nuances of the interventions delivered;
  - define the specific (geocoded) locations of investments;
- Construct an accessibility framework:
  - identify a convenient spatial unit within which to operate (e.g. wards);
  - define an origin-destination matrix for each LEP, including across LEP boundaries if investments are close to the boundary edge[28];
  - apply a routing algorithm to the ODM for a base year and period prior to the investment and annually thereafter to provide an annual record of journey times from any origin to a given destination and taking into account reliability[29];
  - define travel times for a base year and period prior to the investment and annually thereafter;
  - construct an accessibility measure for each spatial unit (ward) included in the analysis[30];
  - construct a variant of the latter with employment fixed at a point prior to investment to act as an instrument;
- Construct an impact geography:
  - map investment locations to OD accessibility profile;
  - define rings/bands of different distances from the site of each investment;
- Cross-reference with the BSD/ARD and extract:
  - details of performance (turnover, employment, productivity, firm numbers) of each business pre/post intervention within each of the control rings;
- Specify estimation model and assess impact:
  - define the nature of the reduced form model that is to be estimated across rings;
  - taking account of any controls/variables believed likely to impact on performance profiles;
  - using fixed effects/differencing and IV techniques.

---

[28] There may be an issue of rail related investments in neighbouring LEPs falling within another LEP matrix if working across boundaries but it is not considered here.

[29] There are differing levels of complexity here regarding measures of reliability and timeframes to be employed in any calculation.

[30] Ultimately, this may produce a DfT accessibility type measure but referencing wards rather than employment/service centres.

Potential alternative and supplementary approaches

LEPs could be asked to propose stations and routes for which data on usage may be either useful for comparative purposes or of interest because it may be indicative of displacement, particularly of retail expenditure. Such comparative data may also need to provide a major element of the evidence base for assessing the effects of the projects involved in transport terms. This needs to include an assessment of whether there is a material impact on congestion/journey times by other modes in the corridors benefitting from the investment.

More important, monitoring data on land values, rentals and physical development in relation to office and retail premises will be a potentially important complement or alternative to VOA data as a basis for assessing local impacts since, as indicated, the latter is unlikely to be sufficiently fine grained for the purpose. Again, LEPs could be asked to propose – and perhaps to collect analogous data in relation to – areas around unimproved stations within their area which could provide useful comparators against which the gross impacts of improvements could be assessed.

The presumption is that rail improvements will have no discernible effects on industrial development. Local development effects in relation to retail activity are clearly of potential interest and may make a notable contribution to local regeneration objectives. However, at a wider level the presumption is that they are entirely displacement so that the only net impacts associated with retail development comprise (any) productivity effects in the centres benefitting from improvements which are identified through the spatial modelling. Similar considerations apply in the case of office development unless the expansions involved form part of wider company developments involving increased export sales (and there is probably only a net impact then if there are constraints on the expansion of the office activities elsewhere).

It is, of course, possible that rail improvements will have wider effects on the distribution of office activities. However, except to the extent that they facilitate inward investment by internationally mobile businesses which would be unwilling to consider alternative locations (likely only in relation to London); it is unlikely that this would have significant impacts at a wider national level.

An overall economic appraisal will, again, require consideration of the transport efficiency benefits of the schemes concerned, ideally based upon multi-modal modelling where the necessary tools and data are available. The suspicion is that systematic assessment may well be challenging in relation to many station improvements.

The assumption is that issue of double counting between local productivity impacts and transport efficiency benefits will not be a major concern in relation to rail schemes since a high proportion of business users are likely to be drawn from outside the immediate area involved.

Summary and conclusions – Rail Transport

An analogous approach to that proposed in relation to other transport schemes based upon estimating the effects on journey time savings by rail between relevant o/d pairs, feeding the findings as an instrument into a fixed effects modelling framework and assessing the extent to which projects appear to improve the performance of locations in the vicinity of the stations involved relative to those in progressively more distant 'rings' is suggested. This is subject to similar caveats about the novelty of this application of the method, the uncertainty about whether many of the investments involved are likely to generate impacts of sufficient scale to be identifiable with any confidence and the weaknesses of some of the relevant datasets.

Again, it is therefore suggested that the LEPs should be expected to undertake supporting monitoring and evaluation activity through assembling data on changes in rentals and on development activity around the stations involved and potential comparators. Similarly, evidence on transport efficiency aspects will be required to support the economic evaluation.

Estimating Gross Impacts: Skills Capital

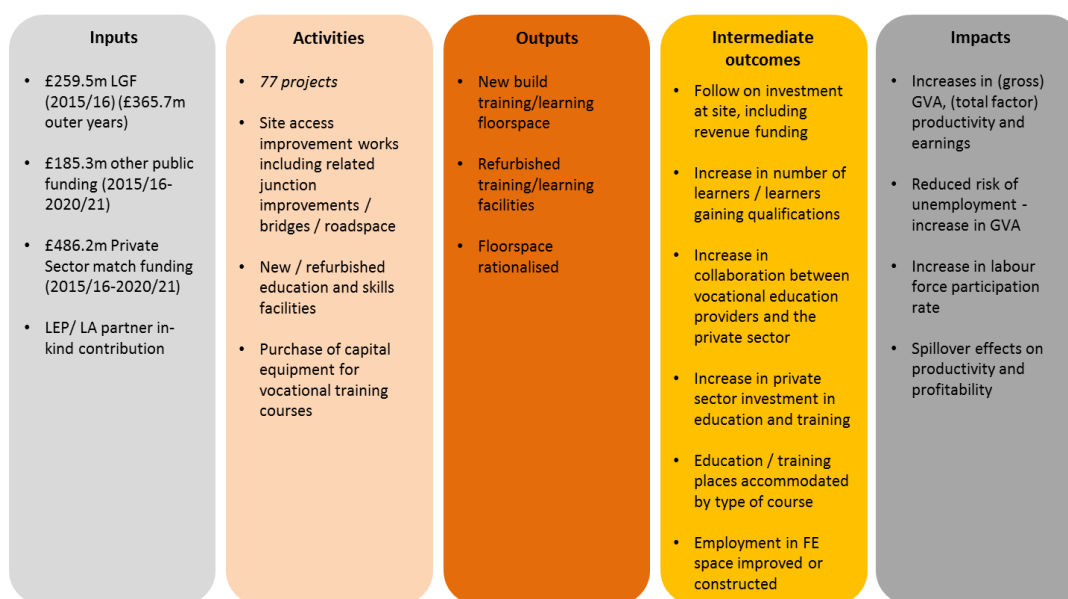## Introduction

Scope of interventions

Growth Deal investment of £625m in 77 projects to increase and improve capacity for education and training is planned between 2015/16 and 2020/21. These projects comprise a mixture of new and refurbished development of further education colleges.

The overwhelming majority of projects concern the physical construction of facilities, such as hub and training centres, further education buildings, and laboratories. The projects vary in size and scope, many smaller initiatives relate only to building modernisation such as classroom refurbishment, whilst others concern entire facility construction. There are also several projects that offer funds specifically for high tech/ value manufacturing capital equipment. LGF funds are also made available for projects, specifically to help with course design, in subjects that target key local growth sectors. These funds support a mixture of interventions aimed at the delivery of skills and enhanced ability to meet local labour market demands. Thus, projects look to increase skills and develop strong businesses and supply chains to boost local economic growth.

Logic chain and monitoring indicators

A logic chain describing the effects of investment in capacity for education and skills is summarised below.

Figure 6: Logic model for skills capital

Inputs	Activities	Outputs	Intermediate outcomes	Impacts
• £259.5m LGF (2015/16) (£365.7m outer years)	• 77 projects	• New build training/learning floorspace	• Follow on investment at site, including revenue funding	• Increases in (gross) GVA, (total factor) productivity and earnings
• £185.3m other public funding (2015/16-2020/21)	• Site access improvement works including related junction improvements / bridges / roadspace	• Refurbished training/learning facilities	• Increase in number of learners / learners gaining qualifications	• Reduced risk of unemployment - increase in GVA
• £486.2m Private Sector match funding (2015/16-2020/21)	• New / refurbished education and skills facilities	• Floorspace rationalised	• Increase in collaboration between vocational education providers and the private sector	• Increase in labour force participation rate
• LEP/ LA partner in-kind contribution	• Purchase of capital equipment for vocational training courses		• Increase in private sector investment in education and training	• Spillover effects on productivity and profitability
			• Education / training places accommodated by type of course	
			• Employment in FE space improved or constructed	

Based on this logic model and the suggested monitoring indicators proposed by BEIS, relevant metrics for data collection and analysis are summarised in the table below.

Table 4: Suggested monitoring indicators for investment in skills capital

OUTPUTS

Measures	Metrics / Units	Potential data sources
New build training / learning floorspace	Sq m	Monitoring information
Refurbished training / learning facilities	Sq m	Monitoring information
Floorspace rationalised	Sq m	Monitoring information

INTERMEDIATE OUTCOMES

Measures	Metrics / Units	Potential data sources
Follow on investment at site, including revenue funding	£, by source	Learning provider/ SFA
Increase in number of learners / learners gaining qualifications	Number by type of qualification	Individualised Learner Records (ILR) https://www.gov.uk/government/collections/individualised-learner-record-ilr
Increase in collaboration between vocational education providers and the private sector	Number of partnerships / value (£m) of sponsorships	Learning provider/ primary research
Increase in private sector investment in education and training	£, by source	Learning provider/ primary research
Education / training places accommodated by type of course	Number by type of course	Learning provider
Employment in FE space improved or constructed	Permanent paid full time equivalent jobs that are directly connected to the	Learning provider

	intervention (#)	

Substantial statistical information is available from the Skills Funding Agency (SFA) regarding the performance of FE colleges. The Individualised Learner Record (ILR) returns that each college is required to submit to the SFA will be particularly helpful in terms of tracking the outcomes of individual learners (although it is commonly known that the ILR has a number of deficiencies, including incomplete records for individual learners).

The extent of private sector support may require the assistance/ permission of the SFA.

Data on other outcome measures is likely to be collated by the learning providers and/ or through primary research.

Potential quasi-experimental and other modelling options

Options and associated issues

There is very limited evidence of quasi-experimental approaches to capital investment in a college environment. The UK studies that do exist are primarily the Frontier Economics and BMG reviews sponsored by BIS. We are guided in our thinking by the experience detailed in these studies.

The bulk of activity consists of extending and improving capacity through new and refurbished development of colleges. There is considerable variety in the form of intervention ranging from relatively limited refurbishment to the construction of buildings, training centres, and laboratories with some provision for capital equipment.

A range of potential metrics are of interest here:

- college activity/performance indicators (numbers participating and rates of participation, success, achievement, retention) as examined in the Frontier Economics/BMG studies;
- medium/long term learner outcomes (additional education, labour market participation);
- wider performance gains to business (turnover, productivity)

This is a broader set of metrics than in other intervention categories. As supported projects are appraised prior to approval, there is the potential for selection bias which may relate to views about recent/future prospects, performance or a range of other decision criteria.

Assessment: Feasibility

In terms of college level activity/performance there are a number of options available including include RDD, matched control and dose-response designs. Individual learner outcomes are more complex in the sense that some form of longitudinal tracking is required in order to follow supported and appropriate non-supported graduates. Potential wider impacts are equally challenging as impacts through improved workforce skills can be

spread far and wide and it is suggested that the approach to these aspects should be based upon wider research evidence.

College level activity/performance

(1) Regression Discontinuity Design (RDD)

In principle, an RDD approach should be feasible to assess college level investment. If project applications are approved/scored according to a common process, details of both successful and unsuccessful applications are available and there exists a threshold for approval that is quasi-random, then there exists (subject to confirmation of other criteria) a prima-facie basis for RDD.

(2) Matching Control

The Frontier Economics/BMG studies, in attempting to assess the general impact of capital investment, declined to consider a matched control study on the basis that most colleges undertook some form of capital spend leaving only a small number of colleges to act as a control group.

The same logic does not necessarily follow as not all colleges will receive support and those colleges that do may not necessarily receive support at the same time. This provides an opportunity to examine a matched control design with matching based on a series of college level and area based observables using PSM or colleges that are selected for assistance in later rounds of support.

(3) Dose-Response Models

Given that there is substantial variation in support levels across projects, and the policy variable is continuous in nature, it is feasible to consider whether a dose-response structure would be suited to the analysis of impact.

Learner Outcomes

(4) Matched Control

The primary assessment vehicle for learner outcomes is likely to be a matched control approach selecting control group members on the basis of a range of individual, area and college characteristics from non-assisted colleges using PSM.

Wider Performance Gains

There are limited prospects through which to assess the impact of capital investment in college facilities on business performance. There is no current basis for linking individuals to employers (other than via survey methods) and thereby no basis on which to assess the scale of any local contribution. Nevertheless, there do exist other BIS studies that may provide a basis for estimating the various impacts of below and above Level 2 learning impacts (see below).

Assessment: Robustness

College level activity/performance

(1) Regression Discontinuity Design (RDD)

As noted, an RDD approach to assess college level investment should be feasible. However, requirements for effective implementation are quite demanding and we suspect that it may be difficult to make the case that they have been met.

(2) Matching Control

There is an opportunity to examine a matched control design with matching based on a series of college level and area based observables using PSM or colleges that are selected for assistance in later rounds of support.

The option involves selecting (as control groups) comparable colleges not subject to intervention and examining differential performance across performance metrics. These would preferably be colleges of similar scale/type serving similar areas and displaying a range of comparable characteristics – which may include some measure of previous capital support/spend.

With PSM providing a basis for selection on observables, there is scope to move beyond the Frontier Economics/BMG approach in terms of techniques. The latter adopted a standard regression approach to performance assessment though it is clear that some of the performance indicators are integers or rates and might be examined in different frameworks[31]. In particular application of Poisson (to counts) and Logit/Probit (to rates) approaches to some of the indicators may provide a more appropriate basis for analysis. With panel data techniques available to model both count rate data and fixed effects and DiD methods used to address time-invariant unobservables, it is likely that this option will achieve a level 3 SMS score.

(3) Dose-Response Models

Advances in dose-response methods make it possible to consider this approach. As outlined elsewhere, advanced techniques are emerging which broaden application of these models but application of such approaches is at an early stage and has not, to our knowledge, been tested in a comparable environment.

Learner Outcomes

(4) Matched Control

This is a demanding option in that the only way to gather information at individual level is via tracking studies that follow (presumably a sample) of individuals in supported colleges. This group would have to be matched with students (presumably via ILR) in non-supported colleges, tested for comparability prior to any analysis and tracking studies undertaken for

---

[31] There are issues about the applicability of ordinary least squares (OLS) to integer and rate data.

the control group also. There is no technical reason why such an approach should not provide robust but we suspect that the practicalities of implementation will be substantial[32].

Wider Performance Gains

We have already noted that there are limited prospects through which to assess the impact of capital investment in college facilities on business performance and there might have to be reliance on broader research.

Assessment: Overview & Implementation

The feasibility and robustness commentary indicates that priority should be initially be given to option (2), perhaps with consideration of option (3) as a longer term possibility. The relevant techniques are well established and rehearsed and should present little in the way of difficulty. In terms of implementation, it will be necessary for evaluators to:

- Liaise with colleges in order to:
    - o understand the nature and nuances of the interventions delivered;
- Implement matching regime:
    - o define college/area/other characteristics from which college control group will be selected;
    - o apply PSM using a range of algorithms assessing balancing and common support characteristics;
    - o identify control group members;
    - o combine treatment and control groups;
    - o assemble longitudinal database of performance indicators for assisted/control group;
- Specify estimation model and assess impact:
    - o define the nature of the reduced form that is to be estimated
    - o apply count/rate models as appropriate;
    - o taking account of any controls/variables believed likely to impact on performance profiles;
    - o using fixed effects techniques;
    - o employing DiD methods

Potential alternative and supplementary approaches

In this case the presumption is that a modelling approach will prove viable but three issues need to be considered:

i) The Frontier Economics/BMG research study identified weaknesses in the ILR participation data and undertook a census of the FE colleges to collect further data on participation, numbers of apprentices trained, the ability of colleges to generate fee income, value of capital expenditure and dates of building work. It would be prudent to assume that the datasets from this earlier work may not be available and that there may

---

[32] It is worth noting that the Frontier Economics/BMG study identified problems with ILR data and included a Census of colleges to fill this and other data gaps

be continuing issues with some of the ILR data. Particularly given the importance of this intervention category within total expenditure, it appears that similar survey work will be required as part of the evaluation.

ii) The previous research found it impracticable to model the impacts of refurbishment expenditure on educational outcomes and explored this and a range of other issues through the inclusion of case studies. Modelling the impacts of this type of expenditure might prove less problematic if one of the alternative approaches discussed above were adopted. However, dependant on the budget which is available for the evaluation, there is clearly an argument for the inclusion of some case studies to explore this and a range of qualitative aspects of the impacts of a sample of projects.

iii) Further evidence will be required to convert findings of impacts on learner numbers/learning outcomes into an assessment of economic impacts. Two BIS studies potentially provide this evidence – although, of course, some updating may be required given the inevitable passage of time until the evaluation is conducted:

a) BIS Research Paper 38 (Cambridge Econometrics and Warwick Institute for Employment Research, 2011) provides a meta-evaluation of the lifetime NPVs of a range of post-19 FE learning (mainly at Levels 2 and 3) taking account of future benefits in terms of higher earnings, better employment prospects and 'spillovers' to other individuals and employers, as well as costs to Government and learners and output foregone during training. Some at least of these estimates are derived from comparison group studies. It is not entirely clear how far costs to Government include college capital costs and it should be noted that the evidence in relation to spillovers on which the study was able to draw is very limited.

b) BIS Research Paper Number 150 (BMG Research and Institute for Employment Studies, 2013) adopts a modelling approach to estimate the impacts of learning below Level 2.

Summary and conclusions – Skills

The proposed approach here is in the first instance to adopt a matching comparator design, contrasting performance of assisted and non-assisted colleges, taking into account a variety of other attributes including previous receipt of capital support. In the event that this design is not feasible, it is proposed to follow previous BIS studies using a longitudinal panel modelling framework to assess the impacts of investments on student numbers and other aspects of attainment, at beneficiary colleges. In either case evidence from BIS sponsored studies based upon a mix of primary research and literature reviews can then be used to assess the likely consequential economic impacts of the identified effects on learner numbers and attainments.

It needs to be noted that the previous studies found it necessary to collect primary data from colleges for the analysis, partly at least in order to address identified weaknesses in some key datasets, particularly the Individual Learner Records. In addition, the passage of time will clearly mean that some of the wider evidence for the impact assessment is likely to need to be 'refreshed' if this is not done as part of other studies for the Department.

Estimating Gross Impacts: Site Development

Site Development – Employment

Scope of interventions

Growth Deal investment of £601m in 95 projects to develop sites for employment and mixed use development is planned between 2015/16 and 2020/21.

The projects funded are primarily concerned with transport/ access related improvement works that have the potential to unlock employment and/ or housing developments. Indeed, around two thirds of the projects relate to transport/ access related improvement works, including junction improvements, capacity/ congestion improvements (e.g. dualling), new access/ relief/ link roads and improvement works on bridges.

There are also a reasonable number of projects that are concerned with site remediation works. Similarly, it includes a small number of mixed use regeneration projects.

Logic chain and monitoring indicators

A logic chain describing the effects of investment in site development for employment and mixed uses is summarised below.

Figure 7: Logic model for employment and mixed-use development
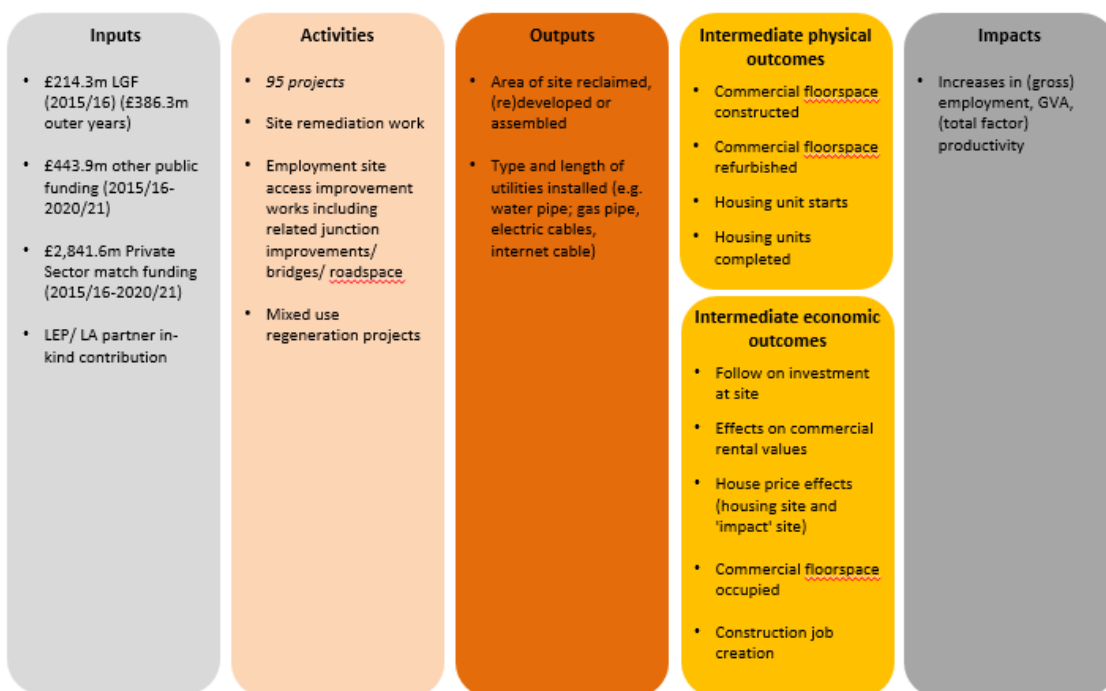
Based on this logic model and the suggested monitoring indicators proposed by BEIS, relevant metrics for data collection and analysis are summarised in the table below.

Table 5: Suggested monitoring indicators for investment in sites for employment and mixed use

OUTPUTS		
Measures	**Metrics / Units**	**Potential data sources**
Area of site reclaimed, (re)developed or assembled	Ha	Monitoring information
Type and length of utilities installed (e.g. water pipe; gas pipe, electric cables, internet cable)	Type and km	Monitoring information

INTERMEDIATE PHYSICAL OUTCOMES		
Measures	**Metrics / Units**	**Potential data sources**
Commercial floorspace constructed	sq m, by class	Monitoring information
Commercial floorspace refurbished	sq m, by class	Monitoring information
Housing unit starts	Number and type	Monitoring information
Housing units completed	Number and type	Monitoring information

INTERMEDIATE ECONOMIC OUTCOMES		
Measures	**Metrics / Units**	**Potential data sources**
Follow on investment at site	Employment/ turnover/ investment/ number of plants	Monitoring information/ primary research
Effects on commercial rental values	£/sq m per month, by class	Monitoring information/ primary research
House price effects (housing site and 'impact' site)	Average property price (£k), by type	Land Registry https://www.gov.uk/government/statistical-data-sets/price-

		paid-data-downloads
Commercial floorspace occupied	sq m, by class	Monitoring information/ primary research

Much of the information required to evaluate projects relating to employment and/ or mixed use development site is likely to come from the monitoring data collected by the scheme sponsors, including, most notably, commercial/ industrial floorspace constructed/ refurbished and housing unit starts/ completions.

Potential effects on commercial rental values may be able to be collected from scheme sponsors, but may also require some primary research.

House price effects could be tracked through Land Registry data (based on a defined area of impact), although this would be restricted to completed sales which may be atypical.

It should be possible to estimate the economic activity associated with the site through datalinking of occupier information and BSD records. Occupier information could either be provided by the scheme sponsors or through primary research.

Primary survey work could also be used to facilitate an analysis of productivity changes (based on information to approximate GVA - total employment costs and profits; and/ or turnover and expenditure on intermediate goods and). Alternatively, turnover per employee could be used as a proxy for productivity.

Potential quasi-experimental and other modelling options

Options and associated issues

Interventions focus on works designed to unlock employment and housing sites through junction improvements, capacity improvements, access/relief roads and improvement of bridges with additional activity to remediate and regenerate sites. An important factor for consideration is the fact that the location of interventions is unlikely to be random and may be located in areas likely to perform better/worse regardless of investment or reflect other factors, clouding attribution of impact. Likewise, projects might be selected taking into account local capacity to implement in a desirable timeframe or greater probability of successful delivery.

A range of metrics are of interest here – particularly turnover, employment (available through the BSD) and productivity (available through the ARD). House price data are also available though access may require discussion with third parties (Land Registry, Banks/Building societies) and may incur costs.

Assessment: Feasibility

(1) Regression Discontinuity Design (RDD)

There is some use of RDD approaches in terms of land use and land change (often zoning related) but very limited use in terms of specific development sites. It is not difficult to understand why this might be the case as there will typically be no quasi-random basis for selection of sites – in practice sites will probably be identified on the basis of readiness or expectations of developer interest, will reflect nuances of local land and property markets and a range of other unobservable rationales.

(2) Instrumental Variables (IV)

As indicated above, the selection of development sites is likely to be endogenous. An IV approach may present a technically feasible approach but we have doubts about identification of a strong instrument and do not consider the option further.

(3) Control group approaches

There are potential opportunities to use control group structures for this intervention given that it will be possible to overlay site boundaries on administrative maps and datasets with some degree of precision.

(3a) Control group approaches: Matching locations

The first option involves selecting (as control groups) comparable sites in broadly similar (local) areas not subject to intervention and examining differential performance across impact metrics. These would preferably be sites at similar stages of preparedness and readiness, subject to similar interventions and subject to similar land and property market considerations. This approach is clearly feasible but dependent on ability to define comparable locations, and probably best undertaken on a site-by-site basis, and there are doubts about its likely robustness which are detailed below.

(3b) Control group approaches: Boundary analysis

A second option might be to regard site boundaries as a spatial discontinuity and to examine performance metrics of on-site organisations with corridors/rings surrounding the site acting as a control group, having controlled for variation in observables. This approach can be adopted either on a scheme-by-scheme basis or across all schemes.

(4) Dose-response models (GPS):

Given the limited numbers of schemes, there is no real scope for the deployment of dose-response models as a feasible vehicle for impact assessment.

Assessment: Robustness

The issue of robustness revolves around whether there exists an appropriate and credible identification strategy through which causality can be assured and policy impact defined. In

what follows we review robustness only for those options that are deemed feasible and thereby set to one side options (1), (2) and (4).

(3a) Control group approaches: Matching locations

Matching treatment/control groups through selection of comparable locations is an obvious way of proceeding. The problem with this approach is finding a set of locations that are observationally (if not un-observationally) equivalent:

- identifying areas that are equivalent in terms of the characteristics of the relevant site will be very demanding, not only in terms of physical attributes of the site, state of readiness, developer interest, economic and property market context but it will be impossible to assess comparators in terms of unobservable characteristics.

It may that project proposals have been submitted from areas not far removed from the schemes funded, comparative schemes might have already been identified or that local expertise might assist identification of alternatives but all are subject to the same deficiency.

Due to the difficulties in matching on observable characteristics alone, this approach might well score at level 2 on the SMS scale.

(3b) Control group approaches: Boundary analysis

The second control group alternative – operating at either side of the site boundary – offers a more robust option in that unobservable local shocks are less likely to impact differentially on the treated/control groups and thereby account for some of any identified policy impact[33].

Identifying firms within site boundaries and in surrounding corridors/rings is straightforward although some GIS input may be required to map sites across administrative maps. To the extent that activity consists of new starts, this can be also be addressed by a similar exercise operating through scrutiny of survival rates.

As the impact datasets of interest will be longitudinal in nature, panel methods are available and it is likely that this option will achieve a level 3 SMS score.

Consideration will also have to be given to the issue of any potential supply-chain effects that result from on-site activity (whether relocation/new starts). Not accounting for any such effects will distort estimates of policy impact and may require some primary evidence from sponsors/firms as to the extent of supply-chain activities in the vicinity of the sites.

---

[33] Faggio (2014) provides an example of a way in which this approach can be implemented defining 'treatment intensity' in terms of circles of differing diameters around sites of interest and accounting for the presence of multiple sites in in close proximity: http://www.spatialeconomics.ac.uk/textonly/SERC/publications/download/sercdp0155.pdf

Assessment: Overview & Implementation

The feasibility and robustness commentary indicates that priority should be given to option (3b). The relevant techniques are well established and rehearsed and should present little in the way of difficulty. Our understanding is that this approach is being used in the current Regional Growth Fund evaluation. In terms of implementation, it will be necessary for evaluators to:

- Liaise with sponsors in order to:
    o understand the nature and nuances of the interventions delivered;
    o define site boundaries;
    o map sites to administrative boundary maps;
    o map a series of boundary corridors/rings, of different distance from the site boundary against administrative boundary maps;
- Cross-reference with the BSD/ARD and extract:
    o details of performance (turnover, employment, productivity) of each business post intervention in/on:
        ▪ sites;
        ▪ corridor/ring areas;
- Specify estimation model and assess impact:
    o define the nature of the reduced form model that is to be estimated across treatment and control areas;
    o taking account of any controls/variables believed likely to impact on performance profiles;
    o using fixed effects/differencing techniques.

Potential alternative and supplementary approaches

The general weakness of spatial modelling approaches in relation to site projects appears to be their likely inability to determine the extent to which development – or different levels of development - have been dependant on the provision of public funding and, if so, whether the level of public funding has been the minimum necessary to secure the development. It is possible to identify benchmark rental figures at which in general terms industrial and office schemes are likely to be potentially commercially viable. However:

- Many of the costs of bringing sites forward for development in terms of the provision of access and services tend to be very specific. In the case of brownfield sites there are also potentially very variable costs of dealing with issues of contamination and dereliction.

- In the case of mixed use schemes commercial viability often depends heavily on the mix of development which is practicable or permissible in planning terms. Particularly in weaker economies, retail and housing uses are often potentially viable whilst industrial and office developments typically requires subsidy.

Our view is that the issues involved could only be assessed through a case study approach. Essentially this would involve utilising a mix of ex-ante and ex-post evidence in a 'residual value' modelling framework comparing the various costs of bring schemes

forward with the potential returns from the sale of plots, with the difference reflecting potential developers' profits or requirements for subsidy if negative.

Whilst a modelling approach ought to be able to allow for local displacement of economic activity, there is clearly a risk that public funding of developments adversely affects the viability and/or timing other developments which might otherwise have come forward. There is clearly a risk of adverse offsetting effects which the modelling approach will not pick up. Case studies looking at the history of development in the locality, site availability and the returns to different types of development have at least some potential to assess the extent of likely displacement of development activity.

As noted, it will potentially helpful for the datalinking analysis and any survey work if the LEP monitoring process identifies the individual firms which are located within the developed sites. As indicated above, survey work may also have a role in relation to the collection of supplementary data to assist the process of matching firms within developments with potential comparators and/or for the assessment of displacement and total productivity effects.

Summary and conclusions – Employment

The projects being funded here appear to have strong similarities to those being funded under the Regional Growth Fund and it clearly makes sense to use the approach which was proposed in the Scoping Study for the programme which, in turn, is firmly based in the recent academic literature. Essentially this involves assessing gross impacts by using areas within concentric rings of increasing distance from the subject sites as comparators, including within the analysis as controls other potential variables which might contribute to observed performance differences.

Datalinking via the ONS Virtual Microdata Laboratory (VML) can usefully be incorporated to explore the extent to which occupiers are new firms or relocations and, if the latter, whether the relocation involved growth – helping to inform an assessment of how far the identified gross effects involve local displacement. There is an additional issue in this case of how far the investment was essential to the sites' development and possible effects on the timing or scale of the development involved. Realistically this can only be assessed through a case study approach given the impracticability of identifying robust comparators.

Site Development – Innovation

Scope of interventions

Growth Deal investment of £126m in 30 projects to develop sites that support business innovation is planned between 2015/16 and 2020/21. These projects comprise the provision of premises for start-ups, spin-outs and existing innovation-related businesses. The overwhelming majority of projects relate to the development of new innovation centre or the expansion/ refurbishment of existing innovation centres.

Logic chain and monitoring indicators

A logic chain describing the effects of investment in site development for innovation is summarised below.

Figure 8: Logic model for site development for innovation uses
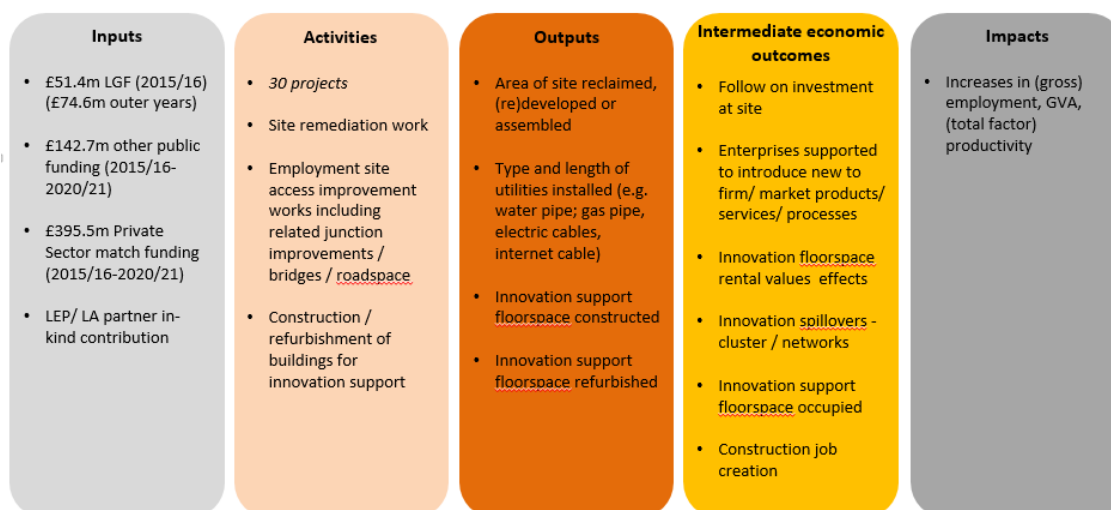
Based on this logic model and the suggested monitoring indicators proposed by BEIS, relevant metrics for data collection and analysis are summarised in the table below.

Table 6: Suggested monitoring indicators for investment in sites supporting innovation

OUTPUTS		
Measures	**Metrics / Units**	**Potential data sources**
Area of site reclaimed, (re)developed or assembled	Ha	Monitoring information
Type and length of utilities installed (e.g. water pipe; gas pipe, electric cables, internet	Type and km	Monitoring information

cable)		
Innovation support floorspace constructed	sq m, by class	Monitoring information
Innovation support floorspace refurbished	sq m, by class	Monitoring information
Start-up units	number	Monitoring information

INTERMEDIATE ECONOMIC OUTCOMES

Measures	Metrics / Units	Potential data sources
Follow on investment at site	Employment/ turnover/ investment/ number of plants	Monitoring information
Effects on innovation floorspace rental values	£/sq m per month, by class	Monitoring information
Innovation spillovers - cluster / networks	Number of participating businesses	Monitoring information/ primary research
Enterprises supported to introduce new to firm/ market products/ services/ processes	Number of enterprises	Monitoring information
Innovation support floorspace occupied	sq m, by class	Monitoring information
Construction job creation	Employment	Primary research/ typical benchmarks

Similar to employment and mixed development sites, much of the information required to evaluate projects relating to innovation sites is likely to come from the monitoring data collected by the scheme sponsors, including, most notably, innovation floorspace constructed/ refurbished.

To facilitate more robust evaluation approaches, the monitoring arrangements proposed by BEIS also require scheme sponsors to supply the following information:

- Number of enterprises assisted to cooperate with research entities/ institutions - The number of treated SMEs working jointly with research entities after assistance. This will be collected on an annual basis for up to three years following the support.

- Number of enterprises supported to introduce new to the market products - The number of treated SMEs that successfully introduce a new-to-market product after assistance. Product should be available for commercial purchase. This will be collected on an annual basis for up to three years following the support.

- Number of enterprises supported to introduce new to the firm products - The number of treated SMEs that successfully introduce a new-to-firm product after assistance. Product should be available for commercial purchase. This will be collected on an annual basis for up to three years following the support.

The above monitoring information would be substantially enhanced if it could also identify and provide basic details (name, address, post code and company reference number (CRN)) on the firms introducing new products and services. This would facilitate datalinking with the BSD and thus allow longitudinal tracking of employment and turnover.

If this is not possible, then primary research is likely to be required. Data on the innovation activity and performance of occupiers and/ or beneficiaries is not be available from secondary data sources.

Potential effects on innovation floorspace rental values should be collected from scheme sponsors, but may also require from primary research.

It should be possible to estimate the economic activity associated with the site by linking occupier information with the BSD. Occupier information could either be provided by the scheme sponsors or through primary research.

Data linking could also be used to measure changes in productivity, although the data required is only available for large firms. Alternatively, turnover per employee (available from the BSD) can be used as a proxy for productivity. Productivity data may also be collected through primary research (i.e. through information on total employment costs and profits; and/ or turnover and expenditure on intermediate goods and services to approximate GVA). It is, however, more challenging to gather robust information from surveys than administrative information.

Potential quasi-experimental and other modelling options

Options and associated issues

Projects in this intervention category are intended to support innovation related start-ups, spin-outs and growth of existing innovation organisations. The bulk of activity takes the form of constructing/refurbishing new or existing innovation centres. As elsewhere, there is an ever present danger of selection bias in the selection of projects to be supported which may relate to views about recent/future prospects or a range of other decision criteria.

The metrics of interest here are primarily sourced through the BSD (turnover, employment) and the ARD (productivity). Given the nature of the intervention, there is also scope for examining survival of new starts which can also be sourced through the BSD.

Assessment: Feasibility

(1) Regression Discontinuity Design (RDD)

RDD approaches are, subject to meeting analytical requirements, suitable for assessment of innovation or R&D support to firms but of limited value in terms of examining innovation related capital/site development. The absence of a quasi-random basis for selection of sites, with sites/centres likely to be identified on the basis of readiness or expectations of developer interest, nuances of local land and property markets and a range of other unobservable rationales, precludes a viable RDD approach.

(2) Instrumental Variables (IV)

As indicated above, the selection of development is likely to be endogenous. An IV approach may present a technically feasible approach but (in common with the review of employment sites) we have strong doubts about identification of a strong instrument and do not consider the option further.

(3) Control group approaches

There are potential opportunities to use control group structures for this intervention given that it will be possible to overlay site boundaries on administrative maps and datasets with some degree of precision.

(3a) Control group: Matching locations/centres

The first option involves selecting (as control groups) comparable sites/centres in broadly similar (local) areas not subject to intervention and examining differential performance across impact metrics. These would preferably be sites/centres at similar stages of preparedness and readiness, subject to similar interventions and subject to similar land and property market considerations. This approach is clearly feasible, dependent on ability to define comparable locations/centres, and probably best undertaken on a site-by-site basis, but there are doubts about both its robustness and the likely extent of potential, available comparators.

(3b) Control group approaches: Boundary analysis

A second option might be to regard site boundaries as a spatial discontinuity and to examine performance metrics on-site with corridors/rings surrounding the site acting as a control group. It is possible that the types of firms operating in the new or improved facilities will not be well represented in narrow corridors beyond site boundaries and it may be necessary to adopt wider control rings to identify an adequate comparator base. This approach can be adopted either on a scheme-by-scheme basis or across all schemes

(3c) Control group: Alternative Comparators

The only difference to (3a) relates to the fact that consideration is given to using businesses assisted at a later stage in the timespan of schemes as a control group for early applicants. The rationale is that this control group might better match differences in unobservable characteristics. There is no fundamental reason why this variant should not

be explored alongside option (3a) and again there is impediment to adopting this approach in either 'aggregation format'.

(4) Dose-response models (GPS)

Given the limited numbers of schemes, there is no real scope for the deployment of dose-response models as a feasible vehicle for impact assessment.

Assessment: Robustness

The issue of robustness revolves around whether there exists an appropriate and credible identification strategy through which causality can be assured and policy impact defined. In what follows we review robustness only for those options that are deemed feasible and thereby set to one side options (1), (2) and (4)[34].

(3a) Control group approaches: Matching locations

Similar to the review of employment sites, the problem with this approach is finding a set of locations that are observationally (if not un-observationally) equivalent:

- identifying sites/centres that are equivalent in terms of the characteristics of the relevant site will be very demanding, not only in terms of physical attributes of the site/centre, state of readiness, developer interest, economic and property market context and it will be impossible to assess comparators in terms of unobservable characteristics.

It may be that project proposals have been submitted from areas not far removed from the schemes funded, comparative schemes might have already been identified or that local expertise might assist identification of alternatives but all are subject to the same deficiency.

Due to the difficulties in matching on observable characteristics alone, this approach might well score at level 2 on the SMS scale.

(3b) Control group approaches: Boundary analysis

The second control group alternative – operating at either side of a boundary- offers a more robust option in that unobservable local shocks are less likely to impact differentially on the treated/control groups and thereby account for some of any identified policy impact. The fact that wider rings may be required does, on the other hand, weaken this argument to a degree.

The process of identifying firms within site boundaries and in surrounding corridors/rings is straightforward although some GIS input may be required to map sites across administrative maps. To the extent that activity consists of new starts, this can be also be addressed by a similar exercise operating through scrutiny of survival rates. As the impact

---

[34] The issue of supply-chain links, as per the discussion in employment sites, also applies here.

datasets of interest are longitudinal in nature and panel methods are available, it is likely that this option will achieve a level 3 SMS score.

(3c) Control group: Alternative Comparators

This option is intrinsically the same as option (3a) but introduces time differences in treatment as a basis improving comparability between treatment and control groups. If it can be demonstrated that late assists, eligible applicants, those of waiting lists are observationally equivalent to early assists, then there is a case to argue that they might well be equivalent in terms of both time variant and invariant unobservables also. In practice, these approaches may be more viable if undertaken across all sites simultaneously in order to boost potential control samples and there may be an opportunity to test matching on a broader range of variables which may make matching more robust. As such, this approach might score 3 or 4 on the SMS scale depending on the quality of the observational match.

Assessment: Overview & Implementation

The feasibility and robustness commentary indicates that priority should be given to option (3c) in the first instance with (3b) as a fall-back. The relevant techniques are well established and rehearsed and should present little in the way of difficulty. In terms of implementation, it will be necessary for evaluators to:

- Liaise with sponsors in order to:
  - understand the nature and nuances of the interventions delivered;
  - define site boundaries;
  - map sites to administrative boundary maps;
- Cross-reference with the BSD/ARD and extract:
  - details of performance (turnover, employment, productivity) of each business post intervention in/on sites;
- Implement matching regime:
  - identify/define lead/lag period or other criteria for alternative assists;
  - collate 'alternative' assists;
  - cross-reference with the BSD/ARD and extract details of performance (turnover, employment, productivity) of each business pre/post intervention
  - test for balancing and common support characteristics;
  - identify 'alternative' control group members;
  - combine treatment and control groups
- Specify estimation model and assess impact:
  - define the nature of the reduced form that is to be estimated
  - taking account of any controls/variables believed likely to impact on performance profiles;
  - using fixed effects techniques;
  - DiD methods.

Potential alternative and supplementary approaches

The issues in this case are discussed above in relation to employment sites. Because of the higher development costs involved the issue of whether schemes might have gone forward in the absence of subsidy – and the associated argument for case studies to explore the possibility of deadweight – is probably less of a concern in relation to Innovation schemes.

Firm based comparisons have the potential to assess the impacts of the innovation support on the scale and success of the innovation activity within the treatment group (in terms of aspects such as R and D spend, patenting activity, Technology Readiness Levels, etc.) and on aspects of business performance. However, it is noted that no information on innovation aspects is likely to be available from secondary sources for most of the businesses concerned so primary survey evidence will be needed if these aspects are to be tracked. Such survey work also has the potential to enhance the range of variables available for matching the treatment group with potential controls.

Summary and conclusions – Innovation

The possibility exists of treating these projects in the same way as general employment schemes to assess their spatial impacts and it is considered that this will be worthwhile. However, the more important evaluation question in this case is the extent to which the provision of innovation support has been effective in encouraging the development of successful innovation based businesses. The likely strong element of (self and policy based) selection bias in the beneficiary group rules out any simple comparison group approach and we believe that the most viable strategies will be the use of:

- Firms coming later to the development of the facilities as controls for those coming earlier based on the assumption that the two groups will share similar characteristics; and/or
- The use of businesses which have sought or are on waiting lists for accommodation as controls based upon similar assumptions.

In either case it will enhance the analysis to incorporate survey work to obtain further details of the business characteristics of the treatment and control groups to improve matching and to provide measures of innovation activity and performance for the two groups which will not typically be available via the VML. Inputs from the LEPs and/or operators of the facilities will be needed to secure contact details for businesses which have sought or are on waiting lists for accommodation and perhaps to help assess their suitability as controls.

# Site Development – Housing

Scope of interventions

Growth Deal investment of £175m in 22 projects to develop housing is planned between 2015/16 and 2020/21. These projects include the provision of social and private housing.

Around half of the housing projects comprise site remediation/ preparation works or land acquisition/ assembly, whilst just under half of the projects are concerned with transport/ access related improvement works that have the potential to unlock housing developments.

Logic chain and monitoring indicators

A logic chain describing the effects of investment in site development for housing is summarised below.

Figure 9: Logic model for housing development
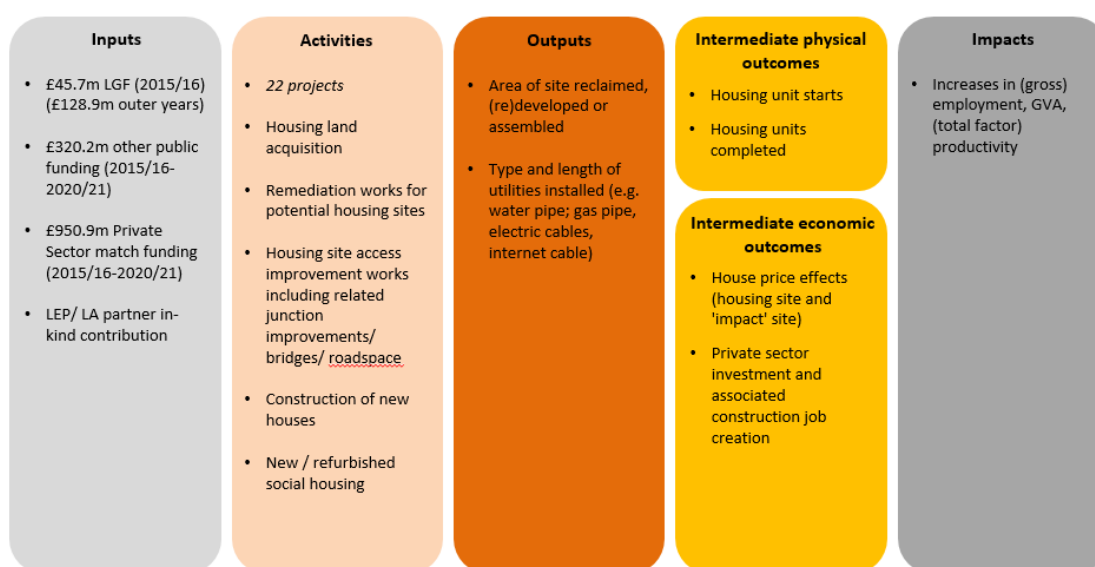
Based on this logic model and the suggested monitoring indicators proposed by BIS, relevant metrics for data collection and analysis are summarised in the table below.

Table 7: Suggested monitoring indicators for investment in housing development

OUTPUTS

Measures	Metrics / Units	Potential data sources
Area of site reclaimed, (re)developed or assembled	Ha	Monitoring information
Type and length of utilities installed (e.g. water pipe; gas	Type and km	Monitoring information

pipe, electric cables, internet cable)		

INTERMEDIATE PHYSICAL OUTCOMES

Measures	Metrics / Units	Potential data sources
Housing unit starts	Number and type	Monitoring information
Housing units completed	Number and type	Monitoring information

INTERMEDIATE ECONOMIC OUTCOMES

Measures	Metrics / Units	Potential data sources
House price effects (housing site and 'impact' site)	Average property price (£k), by type	Land Registry https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads
Private sector investment and associated construction job creation	£m/ employment	Primary research/ monitoring information

Much of the information required to evaluate the housing related site development projects is likely to come from the monitoring data collected by the scheme sponsors, including, most notably, housing unit starts and housing units completed. The monitoring arrangements proposed by BEIS also requires the scheme sponsors to collect data on any follow up investment on the site, if there is a demonstrable link between that investment and the project.

House price effects could be tracked through Land Registry data (based on a defined area of impact), although this would be restricted to completed sales which may be atypical.

Potential quasi-experimental and other modelling options

Options and associated issues

Analysis of projects suggests that the bulk of activity consists of land acquisition and assembly, site/remediation and transport/access designed to unlock sites. This is not that far removed from the profile of activity under the employment/mixed sites intervention category.

As in the latter case, the location of sites is unlikely to be random and will reflect all manner of considerations relating to developer interest, readiness to proceed, planning

restrictions/consents and a range of other factors. Projects might also be selected taking into account local capacity to implement in a desirable timeframe or greater probability of successful delivery.

This intervention is different from others in that primary interest, in relation to impact, lies in terms of the housing market itself.

Assessment: Feasibility

(1) Regression Discontinuity Design (RDD)

It is difficult to believe there will exist a quasi-random basis for selection of sites – in practice sites will probably have been identified on the basis of readiness or expectations of developer interest, will reflect nuances of local land and property markets, planning considerations and a range of other unobservable rationales. This option is set aside.

(2) Instrumental Variables (IV)

The selection of development is likely to be endogenous. An IV approach may present a technically feasible approach but we have doubts about identification of a strong instrument and do not consider the option further.

(3) Control Group approaches

There are potential opportunities to use control group structures for this intervention given that it will be possible to overlay site boundaries on administrative maps and datasets with some degree of precision.

(3a): Control group: Matching locations

The first option involves selecting (as control groups) comparable sites in broadly similar (local) areas not subject to intervention and examining differential performance across housing market metrics. These would preferably be sites at similar stages of preparedness and readiness, subject to similar interventions and subject to similar land and property market considerations. This approach is clearly feasible, dependent on ability to define comparable locations, and probably best undertaken on a site-by-site basis, but there are doubts about robustness which are detailed below.

(3b) Control group approaches: Boundary analysis

Consideration of housing is complicated by the fact that housing development is anticipated not only on 'pure' housing sites but on mixed use sites. It is not clear how this duality might be addressed in terms of an analytical framework.

In terms of the former, assessment of any localised impact can only be examined in terms of whether development can be linked to evidence of change in house market metrics in rings/bands surrounding the site, controlling for variation in observables. This process mimics general treatment intensity approaches and can be adopted either on a scheme-by-scheme basis or across all schemes.

(4) Dose-response models (GPS)

Given the limited numbers of schemes, there is no real scope for the deployment of dose-response models as a feasible vehicle for impact assessment.

Assessment: Robustness

The issue of robustness revolves around whether there exists an appropriate and credible identification strategy through which causality can be assured and policy impact defined. In what follows we review robustness only for those options that are deemed feasible and thereby set to one side options (1), (2) and (4).

(3a) Control group approaches: Matching locations

Matching treatment/control groups through selection of comparable locations is an obvious way of proceeding. The problem with this approach is identifying a set of locations that are equivalent in terms of observational characteristics:

- identifying areas that are equivalent in terms of the characteristics of the relevant site will be very demanding, not only in terms of physical attributes of the site, state of readiness, developer interest, economic and property market context but it will be impossible to assess comparators in terms of unobservable characteristics.

It may that project proposals have been submitted from areas not far removed from the schemes funded, comparative schemes might have already been identified or that local expertise might assist identification of alternatives but all are subject to the same deficiency.

Due to the difficulties in matching on observable characteristics alone, this approach might well score at level 2 on the SMS scale.

(3b) Control group approaches: Boundary analysis

The second control group alternative offers a more robust option in that unobservable local shocks are less likely to impact differentially on the rings and thereby account for some of any identified policy impact. Identifying housing market metrics in surrounding corridors/rings is straightforward although some GIS input may be required to map sites across administrative maps. As the impact datasets of interest will be longitudinal in nature, panel methods will be available and it is feasible that this option will achieve a level 3 SMS score.

Assessment: Overview & Implementation

Analysis of 'pure' housing focussed development is the least well validated element of the scoping exercise. The feasibility and robustness commentary is limited to one option (3b) though even here there is limited evidence of such an approach in practice. If adopted, this option would require evaluators to:

- Liaise with sponsors in order to:
  o understand the nature and nuances of the interventions delivered;

- o define site boundaries;
- o map sites to administrative boundary maps;
- o map a series of boundary corridors/rings, of different distance from the site boundary against administrative boundary maps;
- Cross-reference with the BSD/ARD and extract:
  - o details of performance (turnover, employment, productivity) of each business post intervention in rings surrounding the site.
- Specify estimation model and assess impact:
  - o define the nature of the reduced form model that is to be estimated across ring areas;
  - o taking account of any controls/variables believed likely to impact on performance profiles;
  - o using fixed effects/differencing techniques;

It is our strong view, however, that this is one area that might be better served by consideration of the potential alternatives in the following sub-section.

Potential alternative and supplementary approaches

The presumption is that a modelling approach is not a realistic means of establishing a counterfactual in this case. Housing development is clearly potentially viable in many/most of the areas in which housing schemes are being supported and the need for support arises essentially from planning restrictions limiting development in areas in which the market would be willing to bring forward schemes without subsidy. Arguably therefore the need for public funding arises from a policy rather than a market failure and the counterfactual depends upon how local planning policies would have evolved in the absence of Growth Deal funding – which might have meant no change but which might have involved permitting more development in areas where the market would be willing to develop without subsidy.

The preferred approach in this case is essentially therefore to focus on the assembly of monitoring data on the number of housing units of different types which are provided with Growth Deal funding. Case studies could also have a potentially valuable role, particularly in terms of informing an assessment of the additionality of the development involved.

Summary and conclusions – Housing

The importance of planning policies in shaping the pattern of housing development and the time path of price changes, along with the limited extent to which new housing is likely to drive very local economic change and the likely fairly unique character of many housing sites, mean that any modelling based approach to the definition of counterfactuals will lack credibility. In this case we therefore propose that the evaluation should be based upon a combination of:

- A set of case studies to explore the extent to which Growth Deal funding was crucial to any housing development on the sites concerned or influenced the scale, form or timing of the development. Where housing formed – or potentially could have formed – a component of a mixed use development, there is clearly an issue of how the availability of Growth Deal funding may have influenced the development mix (though this is most likely to have enabled schemes to incorporate a larger employment component with less housing); and

- Aggregation of the housing outputs of various types associated with the Growth Deal schemes.

Estimating Gross Impacts: Business Support

## Introduction

Scope of interventions

Some £136 million of LGF has been committed to 12 business support projects/ programmes between 2015/16 and 2020/21. The overwhelming majority of these are capital grant based schemes. It is worth noting that four of the projects represent continuations of previous RGF programmes.

Six of the 12 projects funded by the LGF in 2015/16, have the primary aim of increasing local and regional employment. These projects target SMEs with growth aspirations, new market entrants as well as young entrepreneurs.

Improving innovation, productivity and skills in local SMEs is another common theme underpinning the projects. These five projects use varied delivery mechanisms that include: tailored business support and advice; support to infrastructure; working directly with businesses; and supply chain development and productivity support.

Logic chain and monitoring indicators

A logic chain describing the effects of investment in business support is summarised below.

Figure 10: Logic model for business support
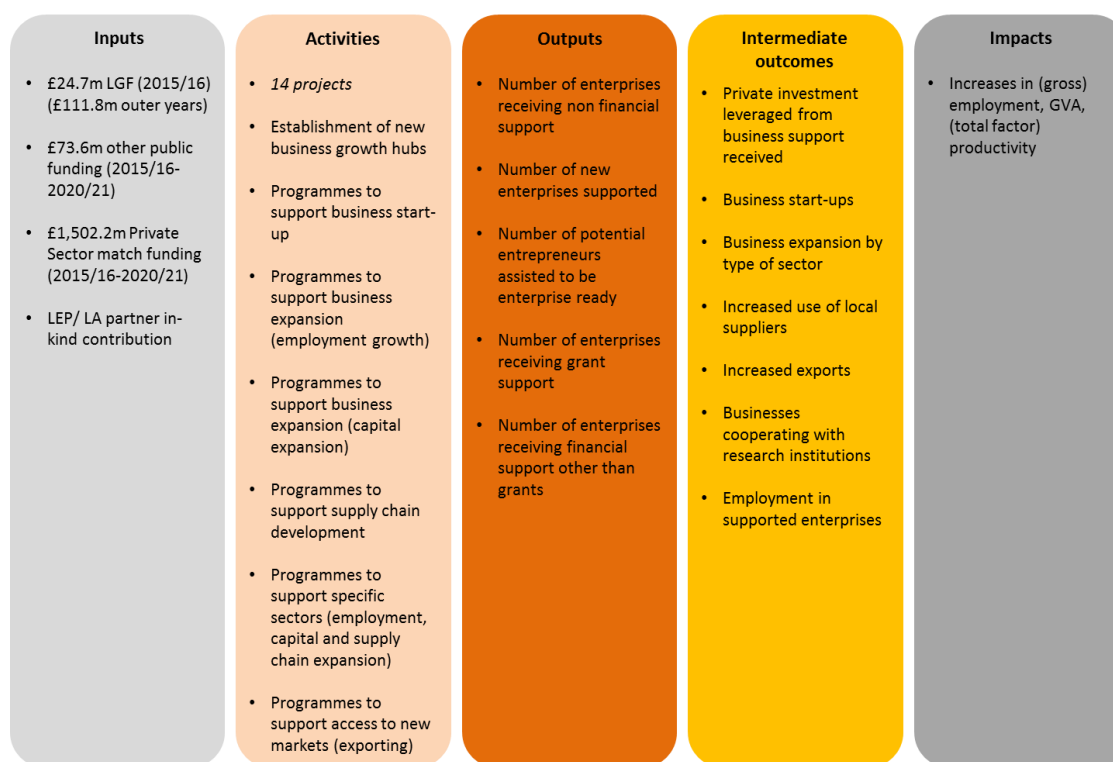
Based on this logic model and the suggested monitoring indicators proposed by BEIS, relevant metrics for data collection and analysis are summarised in the table below.

Table 8: Suggested monitoring indicators for investment in business support

OUTPUTS		
Measures	**Metrics / Units**	**Potential data sources**
Number of enterprises receiving non-financial support	Number, by type of support	Monitoring information
Number of new enterprises supported	Number	Monitoring information
Number of potential entrepreneurs assisted to be enterprise ready	Number	Monitoring information
Number of enterprises receiving grant support	Number	Monitoring information
Number of enterprises receiving financial support	Number	Monitoring information

| other than grits | | |

INTERMEIDATE OUTCOMES

Measures	Metrics / Units	Potential data sources
Private investment leveraged from business support received	£	Primary research
Business start-ups	Number of businesses / value of sales (£m)	BSD http://www.ons.gov.uk/ons/about-ons/business-transparency/freedom-of-information/what-can-i-request/virtual-microdata-laboratory--vml-/index.html
Business expansion by type of sector	Number of businesses / value of additional sales	BSD/ primary research http://www.ons.gov.uk/ons/about-ons/business-transparency/freedom-of-information/what-can-i-request/virtual-microdata-laboratory--vml-/index.html
Increased use of local suppliers	Purchases (£m) by type of supplier	Primary research
Increased exports	£m	Primary research
Businesses cooperating with research institutions	Number of partnerships / value (£m) of sponsorships	Primary research/ monitoring information
Employment in supported enterprises	FTEs	BSD/ Monitoring information (baseline) http://www.ons.gov.uk/ons/about-ons/business-transparency/freedom-of-information/what-can-i-request/virtual-microdata-laboratory--vml-/index.html

The monitoring information on outputs and outcomes for business support programmes is relatively comprehensive. In addition to the core monitoring metrics, scheme sponsors are expected, where relevant, to collect data on the:

- Number of enterprises receiving non-financial support (by type of support)
- Number of new enterprises supported (i.e. trading for less than three years)
- Number of potential entrepreneurs assisted to be enterprise ready
- Number of enterprises receiving grant support
- Number of enterprises receiving financial support other than grants
- Financial return on access to finance schemes (e.g. revolving/ repayable loan funds)

Further to the above outputs and outcomes, and in order to support more robust evaluation approaches, the monitoring arrangements proposed by BEIS also require scheme sponsors to supply the following information:

- Detail of successful and unsuccessful applicants – This will be collected on an ongoing basis and will cover company name, address, post code and company reference number (CRN), as well as a named contact, telephone number and email address (and consent for being contacted).

- Beneficiary characteristics – This will be collected at the point of initial contact and thus will provide a baseline of the business characteristics, including age (year of business registration / founding year), size (turnover and employment) and sector (SIC 2007 1-digit level or higher).

- Other support provided to applicant firm – Other types of support received by successful applicants; covering the scheme, timing, type and value (£) of support received.

- Number of entrepreneurial readiness assists progressing to full trading – The number of potential entrepreneurs assisted that have subsequently progressed to full trading. This will be collected annually.

- Number of enterprises assisted to cooperate with research entities/ institutions – The number of treated SMEs working jointly with research entities after assistance. This will be collected on an annual basis for up to three years following the support.

- Number of enterprises supported to introduce new to the market products – The number of treated SMEs that successfully introduce a new-to-market product after assistance. Product should be available for commercial purchase. This will be collected on an annual basis for up to three years following the support.

- Number of enterprises supported to introduce new to the firm products – The number of treated SMEs that successfully introduce a new-to-firm product after assistance. Product should be available for commercial purchase. This will be collected on an annual basis for up to three years following the support.

Whilst this monitoring information will go a long way in terms of facilitating the application of more advanced evaluation approaches, it could be further enhanced, as noted previously, by the inclusion of the following amendments/ additions:

- Scores of successful and unsuccessful applicants – Where funding has been allocated - largely or wholly - on the basis of a systematic scoring system it would be helpful if the individual scores are made available for both successful and unsuccessful applicants (for example, to provide a basis for a Regression Discontinuity Design evaluation approach or similar).

- Time period for collecting monitoring information – Some of the indicators are expected to be collected annually for up to three years, although this time period could usefully be extended to reflect the full time period over which impacts can be expected to build up. In practice, this is likely to require at least 3-5 years post-completion monitoring. Where a significant part of the project's impacts is expected to take the form of follow-on private sector development a substantially longer monitoring period may (ideally) be required. Equally, in the case of some large investments it may also be necessary to collect some pre-construction data in order to capture potential 'announcement effects'.

- Longitudinal records of a number of performance metrics (such as employment and turnover[35]) for individual firms (at least those that are VAT/ PAYE registered) will be available through secondary data sources such as the BSD. The BSD also includes valuable information on the SIC sector, legal status, ownership and company start date (and termination date, if relevant).

- Data linking could also be used to measure changes in productivity, although the data required is only available for large firms. Alternatively, turnover per employee (available from the BSD) can be used as a proxy for productivity. Productivity data may also be collected through primary research (i.e. through information on total employment costs and profits; and/ or turnover and expenditure on intermediate goods and services to approximate GVA). It is, however, more challenging to gather robust information from surveys than administrative information.

- For unsuccessful applicants, monitoring information from other programmes that are likely to influence similar outcomes as the Growth Deal interventions will be required to ensure a 'clean' control group.

---

[35] Although this is only available for enterprises and not local units.

Potential quasi-experimental and other modelling options

Options and associated issues

Business support interventions cover a range of activities, many capital grant oriented. They target aspirational SMEs, new starts, young entrepreneurs, are designed to improve innovation, productivity and skills and are to be delivered through a variety of mechanisms.

It is anticipated that impacts will primarily be assessed through the metrics of turnover, employment and, where feasible, productivity. Consideration is also given to whether options are viable where assessment might operate i) at an individual scheme/project level with overall impact defined through aggregation ii) on a single combined scheme/project basis[36].

Assessment: Feasibility

(1) Regression Discontinuity Design (RDD)

Application of an RDD approach to the national evaluation is likely to be problematic. While the approach may prove technically viable at individual scheme level, it is doubtful whether practice will accord with necessary implementation conditions at a national level since it is likely that schemes will operate in different ways with different selection criteria and different thresholds. Both of these elements will compromise the aggregation process regardless of whether aggregating impact from individual schemes or assessing impact across all schemes simultaneously. Accordingly, application of an RDD approach to assessing the national impact of business support is not regarded as feasible either on an individual scheme/project basis or a combined basis

(2) Instrumental Variable (IV)

The alternative quasi-random approach alternative to RDD is the use of instrumental variables (IV). As noted elsewhere in his report, identification of appropriate IVs is challenging and is even more so when operating across a wide variety of activities. Use of IVs does not appear to be pervasive in relation to business support interventions and, in the absence of any candidate IV, the approach is not considered further.

(3) Selection Models

Selection models revolve around modelling the selection process itself. This typically involves a two stage process the first of which defines the characteristics likely to make businesses opt to seek assistance (the selection equation) and the second assesses performance taking into account any evidence of selection bias from the first. There is no fundamental reason why this approach should not be applied to individual schemes or across all schemes subject to meeting appropriate implementation requirements.

---

[36] i.e. aggregation formats that include aggregation across schemes post assessment or combined across schemes pre-assessment .

(4) Control group approaches

Control group matching approaches are eminently feasible and are generally well established in programme evaluation. Two variants of such models are considered:

(4a) Control group: Matching organisations

The first option refers to a scenario where treatment is referenced as a binary variable (treated/not treated) and a matching process using propensity scoring matching (PSM) across a range of observable characteristics is employed to establish the composition and validity of appropriate control groups. There would appear to be no impediment to adopting this approach in either 'aggregation format'.

(4b) Control group: Alternative comparators

The only difference to (4a) relates to the fact that consideration is given to using businesses assisted at a later stage in the timespan of schemes as a control group for early applicants. The rationale is that this control group might better match differences in unobservable characteristics. There is no fundamental reason why this variant should not be explored alongside option (4a) and again there is impediment to adopting this approach in either 'aggregation format'.

(5) Dose-response models

The inclusion of an option for dose-response models is to allow for the possibility that it may be desired to consider treatment in a continuous (amount of support) rather than a binary (assisted/non-assisted) format and that interest lies in assessing the variation in treatment effects from different levels of treatment 'exposure' (non-assisted firms receive zero support). Such models have rarely been used in business support evaluations but methodological developments make this approach increasingly more tractable[37]. That said, there is some uncertainty about the applicability of this technique given its novelty and the fact that is untried. There is no evident reason why this approach should not be adopted in either aggregation format though operating across schemes simultaneously would provide greater heterogeneity in modelling.

Assessment: Robustness

The issue of robustness revolves around whether there exists an appropriate and credible identification strategy through which causality can be assured and policy impact defined. In what follows we review robustness only for those options that are deemed feasible and thereby set to one side options (1) and (2).

(3) Selection models

Selection models have an established pedigree in business support studies. If the selection equation contains an appropriate IV that can credibly account for unobservable (time variant and invariant) bias, then such an approach will score at level 4 on the SMS

---

[37] See Bia and Mattei (2007)

scale. A selection equation based on observables alone will score at level 3. In line with our earlier comments, we anticipate limitations in the identification of an appropriate instrument.

(4a) Control group: Matching models

Care is required in the matter of identification in control group matching. Use of a matching approach (such as PSM), where appropriate balancing and common support conditions can be met and demonstrated, provides a means of claiming that the distribution of observable characteristics are broadly the same between treatment and control groups.

This still leaves the issue of potential differences due to (time-invariant and time-variant) unobservable characteristics. As the BSD is longitudinal in nature, there are methods (fixed-effects and difference-in-differences) that can be employed to address time-invariant unobservables though there is limited scope to deal with time-varying unobservables. In the light of the latter, this approach is moderately robust and is likely to score at level 3 on the SMS scale.

(4b) Control group: Matching models (early/late)

This option is intrinsically the same as option (4a) but introduces time differences in treatment as a basis improving comparability between treatment and control groups. If it can be demonstrated that late assists are observationally equivalent to early assists, then there is a case to argue that they might well be equivalent in terms of both time variant and invariant unobservables also. In addition, there may be an opportunity to test matching on a broader range of variables which may make matching more robust. As such, this approach might score 3 or 4 on the SMS scale depending on the quality of the observational match.

(5) Quasi-experimental Dose-response models

There has been some development in dose-response models over recent years, particularly with regard to the use of generalised propensity score (GPS) methods which have properties similar to those of the binary PSM approach. Applications of relevance remain limited though there are indications that variants of the models can address issues such as reverse causality and accommodate IVs. It is probable that this approach might score 3 or 4 on the SMS scale.

Assessment: Overview & Implementation

Consideration of the feasibility and robustness commentary indicates that priority should be given to the two control group match options discussed. One important consideration in this regard is the fact that if assistance is targeted at new or very small organisations, these may well not be represented in the BSD leaving large gaps in the sourcing of information.

Use of the Business Statistics Database (BSD)

Some care and consideration is required in the use of the BSD for impact assessment.

In the first instance, it is important to recognise that the BSD is a snapshot of the Inter-Departmental Business Register (IDBR) taken in April of each year with performance data relating to the previous financial year. Thus, the 2015 BSD will report on performance in the 2014/15 financial year and any intervention that has an impact over the period April 2015 to March 2016 will be reflected in the 2016 BSD rather than the 2015 BSD. It is necessary to ensure that timings of interventions are matched to the appropriate BSD dataset.

Secondly, the BSD is constantly evolving. As such, variable sets in more recent years are more extensive than in the early years. This is not too much of a difficulty in that many of the additions relate to spatial identifiers that can simply be imputed to earlier datasets but it is wise to reflect upon whether this action is appropriate. A related point is that some data series are stored in different data formats across different BSD files and some 'translation' into consistent formats may be required.

Thirdly, representation in the BSD is based on a specific set of criteria relating to the incorporated or unincorporated status of firms, VAT registration/deregistration and/or the presence of at least one individual registered for PAYE. As such, the BSD will not necessarily 'capture' all economic activity in a given location at any single point in time and many micro businesses may not be included. As such, any trend analysis of area-based summary statistics may reflect variation in composition of the datasets as well as change in performance. Care is required to ensure temporal comparability.

Likewise, as firms may register/deregister for VAT to suit their specific circumstances, owner/directors may arrange their remuneration to pay/not pay PAYE and firms may cease trading, there is inevitably a degree of churn in the composition of the BSD from year to year. It is perfectly feasible for some firms to be present in the BSD in some years and not in others and it is feasible that firms receiving assistance in a given year may benefit from that support but not appear in the BSD until some years later, complicating attribution of policy.

The experience of the project team in using the BSD demonstrates that seeking to define a longitudinally consistent dataset of firms over a period of time can prove difficult. This is an important consideration in any analysis that seeks to match individual firms receiving assistance to the BSD. At present, the matching process undertaken by ONS identifies the number of firms that are presented by researchers and that have a presence within the BSD at some point. It does not indicate the number of firms that have a consistent longitudinal presence. It is feasible that any investigation requiring a longitudinally consistent performance profile of a scheme assisting micro/small firms may be limited to a sample that is 70% to 80% lower than initially anticipated.

In this programme context, an effective evaluation may require sponsors to independently collate information as to the performance of businesses as a condition of support or through mandatory follow-up surveys. Documentation of a similar nature may be required of eligible but unsuccessful applicants (more difficult) in order to ensure control group sample are of sufficient scale.

As noted, the early/late variant provides a potentially more robust framework than the standard control match approach though pursuing this option may have implications for the

timing of the evaluation and will be dependent on schemes continuing in much the same form for a number of years in order to generate adequate group sizes.

Whatever approach is adopted, it will be necessary for evaluators to:

- Liaise with sponsors in order to:
    o understand the nature and nuances of the interventions delivered;
    o access the details of the businesses assisted including type of support, timing of support, length of support;
    o access (if available) the details any businesses deemed eligible for assistance but not ultimately supported (with rationale);
- Cross-reference the assisted businesses with the BSD/ARD and extract:
    o details of performance (turnover, employment, productivity) of each business over years prior to and post intervention support;
    o details of business characteristics that will be used to seek matched - businesses in control groups;
- Implement matching regime (a) standard control group:
    o define spatial reference base from which control group will be selected;
    o apply PSM using a range of algorithms assessing balancing and common support characteristics;
    o identify control group members;
    o combine treatment and control groups
- Implement matching regime (b) early/late:
    o identify/define lead/lag period;
    o collate 'late' assists;
    o test for balancing and common support characteristics;
    o identify 'late' control group members;
    o combine treatment and control groups
- Specify estimation model and assess impact:
    o define the nature of the reduced form that is to be estimated taking account of any controls/variables believed likely to impact on performance profiles;
    o using fixed effects techniques;
    o employing DiD methods.

Potential alternative and supplementary approaches

The challenges here appear to arise from the combination of the diversity of the schemes involved and the technical difficulties in applying some of the evaluation techniques described above which probably precludes their use in most locally based evaluations. There is also an issue of whether the numbers of successful and unsuccessful applications are likely to be sufficient to permit a local individual scheme based approach. LEPs could, of course, undertake traditional 'self-reporting' based evaluations, although there will clearly be issues about the credibility of their findings.

The proposed approach to the national evaluation is based upon established methods and there is little reason to consider alternatives. The main issue – and at this stage uncertainty – relates to the question of the numbers of observations which will be available for the different intervention sub-types.

As noted, consideration needs to be given to the inclusion in the evaluation of supplementary survey work to assist the matching with comparators and to provide data to inform the assessment of displacement and total productivity impacts.

Summary and conclusions – Business support

The most viable strategy for the national evaluation is likely to be the use of later assists as a comparison group for earlier assists. If this proves problematic because of sample sizes or timing issues, an alternative would be a longitudinal panel (fixed effects) approach, potentially including unsuccessful as well as successful applicants to increase numbers of observations with dummy variables reflecting the different types of support involved if numbers are too small to allow separate evaluation of different types of scheme. Either approach will benefit if LEPs are able to provide supporting data on applicant characteristics to improve matching or provide potential control variables.

At individual LEP level there may be potential to utilise RDD type methods, depending on the numbers of observations available.

# Estimating Gross Impacts: Flood Management

Introduction

Scope of interventions

Growth Deal investment of £118m in 16 projects to provide flood management is planned between 2015/16 and 2020/21. The overwhelming majority of these comprise the construction of physical flood defences, with the aim of protecting existing business premises and residential properties and/ or unlocking development sites (housing and employment). A couple of the projects are primarily concerned with strengthening the resilience of road and/ or rail infrastructure.

Logic chain and monitoring indicators

A logic chain describing the effects of investment in flood management is summarised below.

Figure 11: Logic model for flood management
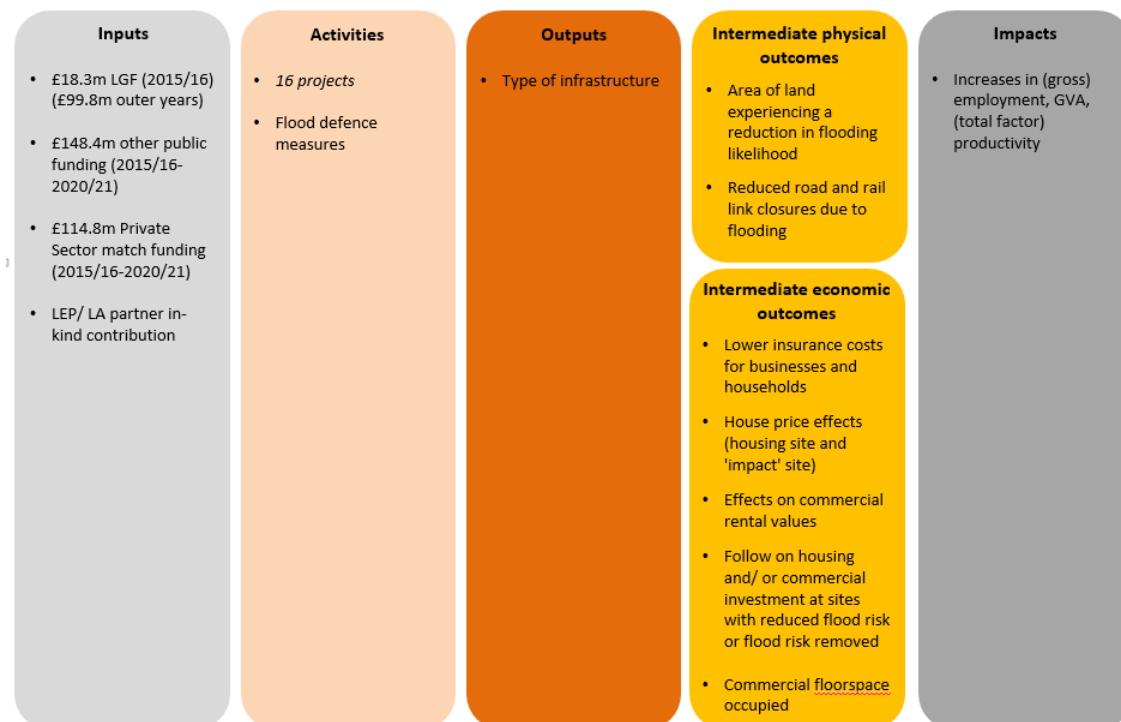
Based on this logic model and the suggested monitoring indicators proposed by BIS, relevant metrics for data collection and analysis are summarised in the table below.

Table 9: Suggested monitoring indicators for investment in flood management

OUTPUTS

Measures	Metrics / Units	Potential data sources
Type of infrastructure	Km, by type	Monitoring information
Area of land experiencing a reduction in flooding likelihood	Ha by land use	Monitoring information/ Environment Agency flood maps http://apps.environment-agency.gov.uk/wiyby/37837.aspx

INTERMEDIATE PHYSICAL OUTCOMES

Measures	Metrics / Units	Potential data sources
Reduced road and rail link closures due to flooding	Number of disruptions by mode per year	Primary research

INTERMEDIATE ECONOMIC OUTCOMES

Measures	Metrics / Units	Potential data sources
Follow on investment at sites	£m, by source	Monitoring information/ primary research
Effects on commercial rental values	£/sq m per month, by class	Primary research/ VOA http://manuals.voa.gov.uk/corporate/publications/statisticsCentralLocalRating.html
House price effects (housing site and 'impact' site)	Average property price (£k), by type	Land Registry https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads
Lower insurance costs for businesses and households	£	Insurance company data/ typical benchmarks/ primary research

Commercial floorspace occupied	sq m, by class	Primary research

The monitoring information collected for flood management investments cover a relatively narrow set of indicators which will need to be complemented by secondary data sources and potentially some primary research.

The monitoring information will be used to determine the type of flood management infrastructure that has been supported and the area of land affected by the investment. This could be complemented by Environment Agency flood maps, which are useful in terms of determining historic/ existing flood risks, as well as identifying other comparable flood risk areas.

Having determined the area of land affected by the investment it will be possible to collect relevant secondary data for relevant business premises, domestic properties (households) and firms located within a best fit geographical area (e.g. based on Super Output Areas or postcodes). To the extent that the investment also unlocks new developments it should also be possible to map these sites against the area of land affected by the investment. Equally, it will be possible to identify roads and rail infrastructure with increased protection from flooding.

It should be feasible to assess the scale and value of economic activity that takes place within the affected area by linking the 'impact' area to BSD information on employment and turnover of firms located within the area.

House price effects on properties located within the affected area could be measured using Land Registry data, although this would be restricted to completed sales which may be atypical.

Identifying actual changes in insurance costs is likely to be very difficult but typical values for insurance cost effects could be applied to the properties and business premises located within the area of impact.

It is unlikely that VOA statistics will be suitable for impact evaluation purposes and as such commercial rental values and floor space occupied effects are likely to require either monitoring information or primary research. Similarly, any information on follow up investment occurring within the affected area, as a result of the investment in flood management measures will need to be collected through monitoring information or primary research.

Potential quasi-experimental and other modelling options

Options and associated issues

All flood management interventions take the form of flood defence measures with primary activities intended to protect business premises or residential properties, road/rail

infrastructure and unlock development sites. A variety of potential impact metrics are of potential interest here:

- metrics readily available include turnover, employment and numbers of firms (all available through the BSD) and productivity (available through the ARD for larger firms);
- house price data are also available through access requires discussion with third parties (Land Registry, Banks/Building societies) and may incur costs;
- floorspace volumes by MSOA used to be published by VOA but now is available only at LAD level and it is not clear whether reporting at the lower level geography has been abandoned or simply removed from publication for convenience/cost;

metrics of interest but not readily available include insurance premia and commercial property rents.

All options reviewed operate on the basis that:

- a prior task – the mapping of flood risk levels to administrative area boundaries – is already available or will be completed prior to any impact assessment and is accessible for areas of proposed investment:

    this will permit the assessment team to locate the site(s) of investment with precision and to define the extent of any new/extended protective footprint in conjunction with project sponsors;

    the adjusted 'risk footprint' associated with investment can be identified in terms of the flood risk/administrative mapping referenced immediately above:

    a time profile of changing risk zones be constructed and mapped against administrative areas for the areas in the vicinity of investments;

Assessment: Feasibility

(1) Regression Discontinuity Design (RDD)

It is possible to argue that flood zone boundaries represent a spatial discontinuity with interventions changing the status of some areas from one flood risk level to another. As such, it may be feasible to employ emerging spatial RDD techniques to examine impact. These approaches are also able to examine the potential for displacement and are probably better applied across all schemes rather than on a scheme-by-scheme basis.

(2) Instrumental Variables (IV)

In common with many policy areas, the location of flood investment may be endogenous[38]. As such, an IV approach may present an effective modus operandi. There also exists a strand of research which is turning to meteorological measures of 'hazard intensity' as the

---

[38] For example, areas of high current/future growth may also be high flood risk rated.

basis of IV approaches in assessing impact of natural disasters[39]. While it is not possible to dismiss IV approaches out of hand, further work would be required to develop and test whether an appropriate IV structure could be determined.

(3) Control group approaches

Flood management presents a number of different opportunities to use control group structures. Given that risk and adjusted risk footprints can be overlaid onto flood risk and administrative maps, controls groups can be defined in very precise spatial terms.

(3a) Control group: Matching locations

The first option involves selecting (as control groups) comparable locations in equivalent risk-level areas not subject to intervention and examining differential performance across impact metrics. These might be areas where similar investments could conceivably be made in the future or could have been made in the past but were not. It may be that:

- proposals were submitted from areas not far removed from the schemes funded that could provide a suitable reference point;
- comparative schemes might already be identified;
- local expertise may be able to define a range of alternatives that might be considered for purposes of comparability.

This approach is feasible though dependent on ability to define comparable locations and the zone of impact. Given the variety in schemes and nature of interventions, it may be more appropriate to apply this option on an individual scheme basis and aggregate to national level.

(3b) Control group: Adjusted risk footprint

A second option might be to define treatment/control areas on the basis of the adjusted risk footprint. Different scenarios are possible in that intervention could introduce:

- a single step reduction in risk within the footprint, with units of interest (e.g. firms, houses) beyond operating as the control group;
- a graduated reduction in risk within the footprint with units of interest beyond operating as the control group.

Both of these dimensions can be accommodated in impact assessment with the second introducing the prospect of multiple 'treatment' levels. There is also the potential to assess differential impacts around the adjusted risk 'boundary' which might assist in identifying the nature and scale of any displacement. This approach can be adopted either on a scheme-by-scheme basis or across all schemes.

---

[39] Miao, Q and Popp, D, (2014), "Necessity as the Mother of Innovation: Innovative responses to natural disasters', Journal or Environmental Economics and Management.

(4) Dose-response models (GPS)

Given the limited numbers of schemes, there is no real scope for the deployment of dose-response models as a feasible vehicle for impact assessment.

Assessment: Robustness

The issue of robustness revolves around whether there exists an appropriate and credible identification strategy through which causality can be assured and policy impact defined. In what follows we review robustness only for those options that are deemed feasible and thereby set to one side option (4).

(1) Regression Discontinuity Design (RDD)

The use of spatial RDD approaches to assess flood management interventions is untried and untested. As such, adopting this methodology would entail a degree of risk. That said, the quasi-random nature of the methodology, if ultimately viewed as credible, would score at level 4 on the SMS scale. We would anticipate that spatial RDD studies will increase in number and, with it, experience of application. This may therefore be an approach that can be exploited more fully at a future stage.

(2) Instrumental Variables (IV)

As noted, use of IV approaches to assessment of environmental and climatic events is expanding. There is, as yet, limited evidence of their deployment in relation to flood events and more detailed consideration of the type of IV that would be suitable would be required prior to adoption of this approach. If the latter provide positive, such an approach would score at level 4 on the SMS scale.

(3a) Control group approaches: Matching locations

Matching treatment/control groups through selection of comparable locations is a simple and inviting strategy. The problem with this approach is finding a set of locations that are observationally (if not un-observationally) equivalent.

Identifying areas that are subject to an equivalent type of risk and equivalent physical, topographical and economic circumstances will be very difficult. The more difficult this becomes, the greater the potential for part of any identified impact to reflect such differences rather than just investment.

As noted earlier, there may be project proposals were submitted from areas not far removed from the schemes funded, comparative schemes that might already be identified or local expertise that might assist identification of alternatives. Unfortunately, all are subject to the same deficiency and due to the difficulties in matching on observable characteristics, this approach might well score at level 2 on the SMS scale.

(3b) Control group: Adjusted risk footprint

The second control group alternative - using the risk footprint or adjusted risk footprint - offers a more robust option. The flood risk map provides a basis for determining both

homogeneity of risk within zones and heterogeneity of risk across zones. At the simplest level, identifying firms within current and adjusted zones is straightforward although i) it may be that some GIS tools may be required to map flood boundaries across administrative maps ii) there may be a (differential) degree of imprecision about current/adjusted boundaries which raises the issue of potential measurement error. There may also be an issue about identifying any previous investments in the areas of interest to ensure that account is taken of these actions in influencing impact metrics.

Beyond this focus, analysing performance of firms within proximity of the risk boundary or adjusted boundary also provides a base for arguing that unobservable local shocks are less likely to impact differentially and account for some of any implied policy impact. As the datasets of interest are longitudinal in nature, fixed-effects methods and DiD techniques can also be employed.

On the basis of these considerations, it is likely that this option will achieve a level 3 SMS score. It may also have the added bonus that the information collated might prove of value in testing a spatial RDD approach at some stage in the evaluation. The identification and addition of a suitable and robust IV would raise the SMS score to level 4.

Assessment: Overview & Implementation

Consideration of the feasibility and robustness commentary indicates that priority should be initially be given to option (3b) with options (1) and (2) open for review at some later stage. The relevant techniques are well established and rehearsed and should present little in the way of difficulty. In terms of implementation, it will be necessary for evaluators to:

- Liaise with sponsors in order to:
    o understand the nature and nuances of the interventions delivered;
    o define the boundaries of current/envisaged flood risk profiles;
    o map the profiles across administrative boundary maps;
    o map a series of boundary 'corridors', of different distance from the revised zone boundary against administrative boundary maps;
- Cross-reference flood risk information with the BSD/ARD and extract:
    o details of performance (turnover, employment, productivity) of each business over years prior to and post intervention support in:
    o relevant flood zones;
    o corridor areas
- Specify estimation model and assess impact:
    o define the nature of the reduced form model that is to be estimated across
        ▪ Treatment/control zones;
        ▪ Treatment/control corridor areas
    o taking account of any controls/variables believed likely to impact on performance profiles;
    o using fixed effects techniques;
    o employing DiD methods.

Potential alternative and supplementary approaches

The proposed modelling methods are judged to represent reasonable approaches to the assessment of economic development and house price effects. Two potential issues for consideration and perhaps additional study are:

a) Effects of reducing flood risks to industrial and commercial property which might be assessed by:

- Assessing numbers of properties of different types within the zones where risks have been removed or reduced and estimating the resultant benefits based upon the typical difference these risk reductions could be expected to make to the insurance premia faced by the businesses involved.

- Determining actual impacts on insurance premia if the necessary data could be accessed – although this seems likely to be problematic.

- Case study research on impacts on local industrial and commercial land values. In the probable absence of systematic VOA data on average values this would likely require some case study research.

b) Consideration of the extent to which house price effects are a satisfactory measure of welfare benefits to residents. It is well known that house price changes have a range of limitations as welfare measures, not least because they will, at best, reflect the valuations of those involved in transactions which may well be atypical.

Given the relatively limited scale of the expenditure involved in such schemes it could clearly be argued that the issues here should be acknowledged but need not be a focus for the evaluation.

Summary and conclusions – Flood management

The approach proposed takes advantage of existing practice in terms of evaluating the impacts of flood events. Just as footprints of flood damage can be identified, footprints of (additional/modified) flood protection afforded to business and homes can be identified as part of Growth Deal investment programmes. Examining the profile of impact metrics as a result of changes in risk levels will provide a basis for assessment.

The key element of the approach is a capacity to identify pre/post intervention patterns of flood risk and, where necessary, graduated change in level of flood risk, on an annual basis if schemes are implemented over a long-period of time. Flood risk will need to be mapped against administrative boundaries to permit datalinking and account will need to be taken of any previous flood-related investment at or close to the location of intervention. It may be that collection of this information will provide the basis of emerging evaluation techniques at a later stage.

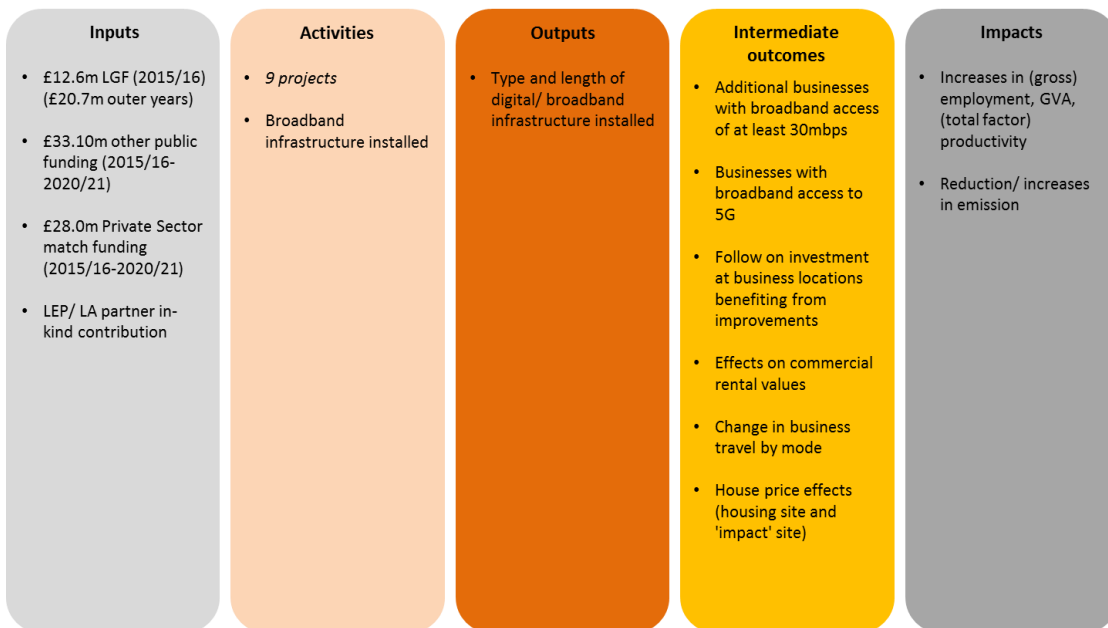# Estimating Gross Impacts: Digital/ Broadband Infrastructure

Introduction

Scope of interventions

Growth Deal investment of £33m in nine projects to provide digital / broadband infrastructure is planned between 2015/16 and 2020/21. The majority of these projects involve investment in both standard and superfast broadband infrastructure, with the aim of speeding up roll out and/ or encouraging broadband roll-out to more rural areas. Three of the projects receive investment aimed at facilitating business application of 5G. Logic chain and monitoring indicators

A logic chain describing the effects of investment in digital and broadband infrastructure is summarised below.

Figure 12: Logic model for digital/ broadband infrastructure

Inputs	Activities	Outputs	Intermediate outcomes	Impacts
• £12.6m LGF (2015/16) (£20.7m outer years) • £33.10m other public funding (2015/16-2020/21) • £28.0m Private Sector match funding (2015/16-2020/21) • LEP/ LA partner in-kind contribution	• *9 projects* • Broadband infrastructure installed	• Type and length of digital/ broadband infrastructure installed	• Additional businesses with broadband access of at least 30mbps • Businesses with broadband access to 5G • Follow on investment at business locations benefiting from improvements • Effects on commercial rental values • Change in business travel by mode • House price effects (housing site and 'impact' site)	• Increases in (gross) employment, GVA, (total factor) productivity • Reduction/ increases in emission

Based on this logic model and the suggested monitoring indicators proposed by BEIS, relevant metrics for data collection and analysis are summarised in the table below.

Table 10: Suggested monitoring indicators for investment in digital / broadband infrastructure

OUTPUTS		
Measures	**Metrics / Units**	**Potential data sources**
Type and length of digital/ broadband infrastructure installed	Type and km	Monitoring information

INTERMEDIATE OUTCOMES		
Measures	**Metrics / Units**	**Potential data sources**
Additional businesses with broadband access of at least 30mbps	Number of businesses connected by sector / size	Ofcom broadband statistics (average speeds by post code) http://maps.ofcom.org.uk/broadband/broadband-data/
Businesses with broadband access to 5G	Number of businesses connected by sector / size	Monitoring information
Follow on investment at site	£, by source	Monitoring information
Effects on commercial rental values	£/m2	Monitoring information/ VOA http://manuals.voa.gov.uk/corporate/publications/statisticsCentralLocalRating.html
Change in business travel by mode	No of pkm by mode	DfT traffic counts http://www.dft.gov.uk/traffic-counts/
Possible house price effects (housing site and 'impact' site)	Average property price (£k), by type	Land Registry https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads

Beyond the core indicators, the monitoring information being collected by scheme sponsors of broadband projects is currently limited to the number of additional commercial premises that, as a result of the intervention, have access to average broadband speeds of at least 30mbps (where this was not previously achieved). It can be argued that this should be extended to include details of the infrastructure installed and the broadband enabled exchanges affected by the investment (although this data may be implicit in the collection of data on the commercial premises affected).

In order to be able to measure any potential house price effects, there is also a case for scheme sponsors to collect information on the number of additional domestic properties that, as a result of the intervention, have access to average broadband speeds of at least 30mbps.

Ideally, this monitoring information should also include some form of geocoding (e.g. the address) of the affected commercial premises and domestic properties. This would allow datalinking with secondary data sources and/ or provide a basis for a survey sampling.

In terms of measuring potential business impacts, there should be opportunities for linking the commercial premises that have benefited from improved broadband speeds and longitudinal BSD data on employment and turnover for affected firms. Similarly, house price effects could be tracked through Land Registry data (based on a defined area of impact), although this would be restricted to completed sales which may be atypical.

Data linking could also be used to measure changes in productivity, although the data required is only available for large firms. Alternatively, turnover per employee (available from the BSD) can be used as a proxy for productivity. Productivity data may also be collected through primary research (i.e. through information on total employment costs and profits; and/ or turnover and expenditure on intermediate goods and services to approximate GVA). It is, however, more challenging to gather robust information from surveys than administrative information.

Potential quasi-experimental and other modelling options

Options and associated issues

Digital infrastructure interventions take the form of expanding coverage of standard and superfast broadband (primarily to rural areas) and facilitating development of the 5G network. In common with other intervention areas, a variety of potential impact metrics are of potential interest here:

- metrics readily available include turnover, employment and numbers of firms (all available through the BSD) and productivity (available through the ARD);
- house price data are also available though access requires discussion with third parties (Land Registry, Banks/Building societies) and may incur costs;
- floorspace volumes by MSOA used to be published by VOA but now is available only at LAD level and it is not clear whether reporting at the lower level geography has been abandoned or simply removed from publication for convenience/cost;

- metrics of interest but not readily available include insurance premiums and commercial property rent.

All options reviewed operate on the basis that:

- a prior task - the mapping of broadband availability/speed to administrative area boundaries - is already available or will be completed prior to any impact assessment and is accessible for areas of proposed investment:

  adjusted coverage associated with investment can be identified in terms of the mapping referenced immediately above:

  a time profile of changing coverage/speed be constructed and mapped against administrative areas for the areas in the vicinity of investments;

Assessment: Feasibility

(1) Regression Discontinuity Design (RDD)

It is possible to argue that coverage/speed boundaries represent a spatial discontinuity with some areas benefitting from new/upgraded technology and infrastructure and investment changing the status of some areas from one technology level to another. As such, it may be feasible to employ emerging spatial RDD techniques to examine impact. These approaches are also able to examine the potential for displacement and are probably better applied across all schemes rather than on a scheme-by-scheme basis.

(2) Instrumental Variables (IV)

The location of publicly supported digital investment is likely to be endogenous[40]. As such, an IV approach may present an effective modus operandi. There are instances of IVs being used in broadband studies (including coverage of vintage communications networks).

(3) Control group approaches

Digital infrastructure presents a number of different opportunities to use control group approaches. Given that coverage/speed profiles can be overlaid onto administrative maps, controls groups can be defined in very precise spatial terms.

(3a) Control group approaches: Matching locations

The first option involves selecting (as control groups) comparable locations in equivalent speed/coverage areas prior to investment and examining differential performance across impact metrics. There are several examples of broadband control group areas being selected on the basis of matching approaches such as PSM and there is no reason why this approach should not prove viable.

---

[40] For example, areas of lagging performance may have limited digital access or the difference between urban and rural access may reflect differential growth/customer profiles.

(3b) Control group approaches: Boundary analysis

The second option is a variant of (3a) and essentially uses the same dataset(s) but contrasts change in impact metrics in either side of a common boundary where speed varies at the boundary itself. This option is most suited to the Local Exchange (LE) level where variation in speed across the LE boundary is easily and robustly defined, and thereby dependent on being able to access relevant boundary information.

 (4) Quasi-experimental Dose-response models (GPS)

Given the limited numbers of schemes, and anticipated difficulty of constructing accurate spend information for upgrading of technology on a small spatial scale, there is no real scope for the deployment of dose-response models as a feasible vehicle for impact assessment.

Assessment: Robustness

The issue of robustness revolves around whether there exists an appropriate and credible identification strategy through which causality can be assured and policy impact defined. In what follows we review robustness only for those options that are deemed feasible and thereby set to one side option (4).

(1) Regression Discontinuity Design (RDD)

The use of spatial RDD approaches to assess digital infrastructure interventions is in its infancy. As such, adopting this methodology would entail a degree of risk. That said, the quasi-random nature of the methodology, if ultimately viewed as credible, would score at level 4 on the SMS scale. We would anticipate that spatial RDD studies will increase in number and, with it, experience of application. This may therefore be an approach that can be exploited more fully at a future stage.

(2) Instrumental Variables (IV)

As noted, there do exist examples of IVs being used in broadband studies (including coverage of vintage communications networks) but it unclear at this point whether these translate into the UK environment. There may also be scope for developing a digital accessibility measure which can be fixed prior to investment and which might act as an instrument but this is, at this point, speculative. Accordingly, further work is required to develop and test whether an appropriate IV structure can be determined. If the latter proves positive, such an approach would score at level 4 on the SMS scale.

(3a) Control group approaches: Matching locations

One of the practical considerations in this option is the spatial level at which a match might be made. Of primary importance here is the basis on which coverage/speed profiles are available.

The most directly relevant spatial unit is that at Local Exchange (LE) level - the points of delivery for broadband services to customers and of which there are close to 4,000 in England[41]. Each LE has a defined capacity and coverage providing an authoritative source of diffusion over time although such information is not publicly available[42]. It is unlikely that LE coverage will map conveniently to administrative datasets, making the matching process somewhat complex, but the extent of any inconsistency is unknown and requires further investigation.

One alternative is the fixed broadband dataset provided by Ofcom which details average, median and maximum speeds along with availability of superfast broadband and number of connections by postcode. Data are suppressed if there are less than three residential or business premises but are now available on an annual basis. Recognising the caveats about suppression, it should be feasible to aggregate this dataset to a convenient spatial basis to be set against performance data. This would generate coverage/speed maps that are longitudinal.

As such, fixed-effects methods and DiD techniques might also be employed though the recent vintage of some variants of broadband related data may require careful assessment of whether there is a sufficient span of pre intervention information. On the basis of these considerations, it is likely that this option will achieve a level 3 SMS score.

(3b) Control group approaches: Boundary analysis

Analysing performance of firms within proximity of a fixed boundary provides a base for arguing that unobservable local shocks are less likely to impact differentially and account for some of any implied policy impact. This is arguably a more robust approach that (3a) but is probably only so in the context of LE level data where boundaries are fixed and robust and would require GIS input. Nevertheless, it is likely that this option will achieve a level 3 SMS score. It may also have the added bonus that the information collated might prove of value in testing a spatial RDD approach at some stage in the evaluation.

Assessment: Overview & Implementation

It is clear that a priority for this intervention involves the construction or delineation of a coverage/speed map that can matched to some level of administrative boundary. Subject to this prerequisite, the feasibility and robustness commentary indicates that option (3a) is the most straightforward to pursue with options (1), (2) and (3b) open for review at some later stage[43]. The relevant techniques are well established and rehearsed and should present little in the way of difficulty. In terms of implementation, it will be necessary for evaluators to:

- Liaise with sponsors in order to:
    o understand the nature and nuances of the interventions delivered;
    o identify the (time and spatial) pattern of roll out;
- Define the accessibility/speed map:

---

[41] The 'local loop'.
[42] The dataset has been made available for purposes of academic study.
[43] This proposed approach is similar in nature to the Ahlfeldt et al (2014) study.

- - o define the spatial unit at which accessibility/speed data is available;
    - o map profiles to administrative boundary datasets;
    - o construct a longitudinal map of accessibility/speed to administrative and control areas;
  - Cross-reference accessibility/speed map with the BSD/ARD and extract:
    - o details of performance (turnover, employment, productivity) of each business over years prior to and post intervention support in:
    - o intervention areas;
    - o control areas
  - Specify estimation model and assess impact:
    - o define the nature of the reduced form model that is to be estimated across treatment/control areas;
    - o taking account of any controls/variables believed likely to impact on performance profiles;
    - o using fixed effects techniques;
    - o employing DiD methods.

Potential alternative and supplementary approaches

Two issues require consideration in this case:

a) The additionality of the investment involved. If there were an effective competitive market in the provision of broadband infrastructure and the investment schemes have been tendered through an OJEU compliant process, as is assumed to be the case, the public sector costs involved should naturally have been minimised. The complication is clearly the existence of a dominant player which is understood to have won at least most of the contracts. The issues involved clearly could be considered as part of the evaluation, perhaps through some case studies of the costs and potential revenues, through benchmarking against the costs of analogous schemes supported through other public programmes or potentially through an assessment of the factors which drive requirements for subsidy based upon the characteristics of areas which have and have not benefitted from the provision of the investment involved through normal market mechanisms.

Clearly there is a question of whether the issues here should be a focus for the Growth Deals evaluation or whether they are a matter for a wider study, particularly bearing in mind that such investments are a relatively small element of the Growth Deals programme.

b) The wider welfare - consumer and producer surplus - benefits of the investment. House price impacts could again potentially be regarded as capitalising the former, albeit imperfectly. The latter could in principle be assessed in terms of impacts on firm profitability, although as the schemes are assumed to be focused on rural areas dominated by small firms for which profits data is not available via the VML, it seems unlikely that this approach will be worth pursuing given the relatively limited scale of the investment involved and the likelihood that such effects are likely to represent a relatively minor part of their total impacts. As an alternative, it may be that useful information will be available from evaluations of analogous projects.

Summary and conclusions

It is proposed that the impact of broadband investment be examined through an approach which bears some similarity to recent academic studies in the area. In effect, the approach operates by contrasting the impact of changing coverage and speed across comparator areas and examining whether there is evidence of differential impact as investment is rolled out.

As with flood management, the key element of the approach is a capacity to identify pre/post intervention patterns of broadband coverage and speed on an annual basis which can be mapped against administrative areas to permit datalinking. The construction of coverage/speed maps is an area of uncertainty though academic studies have been able to access quite detailed information with regard to local capacity and the Ofcom fixed broadband dataset provides capacity/speed information down to postcode level. Some attention may need to be paid to the extent of pre-intervention trends but collection of this information will again also provide a basis for use of emerging evaluation techniques at a later stage.

Economic Evaluation Framework

Introduction

The economic evaluation will need to utilise and build upon the findings of the impact assessment in particular in terms of the employment and GVA which is estimated to have been created over different timescales at national level by the Growth Deals programme. Any limitations or uncertainties in the findings of the impact assessment work will, of course, inevitably carry through directly to the economic evaluation.

The economic evaluation will be focused on the question of how far the programme has represented value for money, including the issues of:

- Economy: in this context, whether projects are being delivered at the minimum cost which is practicably achievable. The main insights to this issue are likely to come from the process evaluation. It will also be useful to include comparisons with typical unit costs for the types of project involved, although comparisons will need to be sensitive to potential project 'abnormals';

- Efficiency: whether the maximum outputs and impacts are being achieved from the resources involved. This aspect will primarily utilise the results of the impact assessment along with evidence in relation to external comparators; and

- Effectiveness: how far the overall objectives of the programme have been achieved, again based largely on the findings from the impact assessment.

In broad terms there are two main options for the economic evaluation:

- A cost-effectiveness analysis based upon a comparison of the total public sector costs of particular interventions or the Growth Deals as a whole with the estimated resulting gross and/or net employment and/or GVA to provide, for example, estimates of cost per gross and net job created.

- A cost-benefit analysis (cba) comparing the net present value (npv) of the costs with the npv of the total welfare benefits generated by the Growth Deals.

The cba approach is clearly strongly to be preferred in general terms as it potentially provides a much more rounded picture of the value for money (vfm) of the expenditure involved. In this context the case for such a wider approach is particularly compelling because many of the intervention types will generate a range of benefits which will not be reflected in their estimated employment and GVA impacts, limiting the usefulness of cost-effectiveness measures.

Costs to be assessed

Either approach will require the estimation of a combination of:

- The 'process' costs to the Government Departments, LEPs, local authorities and other public sector bodies involved of all phases of the development and implementation of the programme. In practice these would probably be very difficult and fairly costly to estimate with any precision and, as they are likely to represent a relatively minor element in the overall costs of the programme, a fairly 'light touch' approach to their estimation – based upon small scale survey work and application of the standard cost model – is likely to be appropriate.

- The total public sector costs of the projects under assessment – irrespective of the funding sources involved, as indicated above.

A cost-benefit analysis will also in principle require the estimation of the net returns to the private sector both of the projects involved and of the follow-on private sector activities to which they lead. In many/most cases the project appraisal process (for example, for grants to businesses or support for site investments) or the tendering process (for example, in relation to broadband projects) should tend to drive the net returns to the private sector down to zero, taking account of risk adjusted capital costs. Assuming that the process evaluation confirms these are reasonable working assumptions, it may therefore be reasonable to exclude any detailed consideration of these aspects in the national evaluation.

Cost-effectiveness analysis

Comparisons of the unit costs of achieving particular impacts could be made across intervention categories or, if the modelling or impact assessments are structured to permit this, across groups of LEPs or projects. Comparisons could also be made with benchmarks from other evaluations though, to be meaningful, the evaluations must have been undertaken with comparable frameworks and levels of rigour. The availability of comparators meeting this criterion is arguably currently very limited.

Such an approach has advantages in terms of simplicity and avoidance of the requirements for much of the supporting analysis to assess wider impacts proposed in the succeeding sub-sections. It can also yield useful insights, particularly in relation to those intervention types which are primarily focused on generating local economic development.

However, as indicated, the approach would not provide an overall assessment of the vfm of the programme. Additionally, any comparisons of vfm between intervention categories would be substantially distorted by the failure to take account of the wider benefits (and perhaps costs) of different types of intervention.

Cost-benefit analysis

Three types of benefit would need to be included in the analysis.

Short term GVA impacts

Following general BEIS practice the assumption is that these will be assessed as the npv of the GVA impacts of the interventions over the period they are estimated to persist. In the short term these are likely to be primarily the result of increases in economic activity and employment.

GVA is arguably not an entirely satisfactory measure of the value of the economic impacts as it does not take into account private sector capital consumption (this would be accounted for in the alternative Net Value Added measure – though seeking to utilise this would involve a range of further complications and data requirements). Nor does it take into account factors such as commuting and childcare costs or opportunity costs in terms of foregone leisure, on the one hand, or the Exchequer savings and the various private and social benefits of reductions in unemployment on the other.

We are aware from other studies that the circumstances under which it is reasonable to include the flow of GVA associated with net additional jobs as a benefit may be the subject of future debate with Treasury.

Long term GVA impacts

Following the approach set out earlier, these need to be assessed as the npv of the GVA impacts associated with effects on productivity beyond the period for which the short term effects via impacts on economic activity are estimated to persist. In the case of skills related effects there is clear research evidence of such long term impacts. It is less clear how far ongoing total factor productivity impacts will be identifiable in relation to other intervention types or, if they are identifiable within the modelling, how long they should be assumed to continue. It may be that this will require a scenario/sensitivity analysis based approach.

Wider benefits

Table 11 provides a summary of the wider potential benefits of the interventions involved and our proposals for dealing with these building upon the approaches set out above.

Table 11: Proposed Approaches to Wider Potential Benefits

Intervention Category	Potential Wider Impacts	Proposed Approaches to Assessment
Transport	Transport efficiency benefits	WebTAG based (ex-ante) assessments will be available for at least some schemes – there is potential to update these using (ex-post) monitoring data, particularly for more major schemes
	Wider economic benefits as per WebTAG	Some evidence available from appraisals or could be generated through modelling of transport efficiency aspects for major schemes. There would be a need adjust for double counting with modelled local productivity

	CO2 impacts	impacts (see above) Some, probably very partial, evidence should be available from DfT monitoring requirements for sustainable development investments. Potential to value based on DECC guidance but unlikely to be a significant aspect
Skills Capital	Spillover benefits to employers and individuals	Estimates available from other BIS research (see above)
Sites	Uplift in land values	Potentially assessable in broad terms in the case of housing land via combination of monitoring and VOA data. Suitable VOA data is not generally available for employment land so any assessment would require primary evidence
	Housing market effects	Effects on general level of house prices, social housing waiting lists, etc are likely to be too marginal to be discernible through modelling
Business Support	None seem to warrant research	
Flood Management	Uplift in land and property values	Potential to model impacts on house prices
	Reduction in insurance premia (likely to be reflected in uplifts above)	Possibilities to assess impacts on industrial/commercial insurance premia (see above)
Digital/Broadband	Welfare benefits - possible capitalisation in house prices	Modelling via house price data (see above)

Summary and conclusions

The preferred approach is to utilise a cost-benefit framework in which:

Costs are assessed as the sum of the net present values of the public sector process/implementation and project costs, irrespective of the source of the funding involved.

The benefits comprise the sum of the net present values of the:

- Short term impacts on GVA, primarily associated with effects on economic activity.
- Long term impacts on GVA associated with productivity related effects
  Wider benefits of the interventions involved.

It is anticipated that there will inevitably be significant gaps in the available evidence which will need to be dealt with through one or a combination of approaches:

- Focussing just on those elements for which sufficient evidence is available.
- Extrapolating the evidence which is available to analogous aspects where insufficient information is available – although it needs to be recognised that there will be limits on the extent to which this will be reasonable.
- Use of scenario/sensitivity analysis in relation to areas where evidence is deficient or non-existent to explore the robustness of the emerging conclusions.

Table 12: Summary of potentially robust statistical approaches to the ex post evaluation of gross outcomes (additionality), by type of intervention

Type of intervention	Quasi-random approaches			Quasi-experimental – Matching strategies			Other approaches	
	Instrumental Variables	**Regression Discontinuity Design**	**Selection models**	**Control group approach**	**Control group – boundary approach**	**Control group – alternative comparators**	**Dose response**	**Accessibility models**
Transport	✓✓ Accessibility indices				✓✓✓ Concentric rings		✓ Generalised propensity scores	✓ Accessibility models
Skills capital – college level		✓ College investment		✓✓ PSM of colleges or Panel Data			✓ Levels of support	
Skills capital – learner outcomes				✓✓ PSM of learners				
Site development – employment / innovation				✓✓ Locations	✓✓✓ Concentric rings	✓✓✓ Early / late		
Site development – housing				✓✓ Locations	✓✓✓ Concentric rings			

Business support			✓ Selection / IV	✓✓✓ PSM / FE of firms		✓✓ Early / late	✓ Generalised propensity scores	
Flood management	✓ Meteorological	✓ Spatial approach		✓✓ Locations	✓✓ Adjusted risk			
Digital / broadband infrastructure	✓ Vintage or accessibility	✓ Spatial approach		✓✓✓ Locations	✓✓ Boundary neighbourhood			✓ Digital accessibility models

Key: ✓ = potential but needs development / testing

✓✓ = developed and previously tested approach but significant challenges to the application in the Growth Deal context

✓✓✓ = developed and previously tested approach, suitable for application in the Growth Deal context

Shaded cells represent the most currently feasible approach for each type of intervention

Conclusions by type of intervention

Transport – Road Improvements

The proposed approach here builds upon recent academic studies. It involves the construction of pre and post-investment matrices of journey times by road to identify potentially significant travel time savings between relevant origin/destination pairs. These would then feed as an instrument, along with relevant control variables, into a fixed effects modelling framework designed to assess the extent to which the Growth Deal investments 'explain' variations in small area economic performance between locations at varying distances from the schemes involved as a basis for assessing their impacts.

There needs to be a recognition that the nature and limited scale of many of the investments involved and the potential issues in incorporating effects in alleviating congestion create a particularly challenging context for the application of this method. For this reason and to provide a richer evidence base with which the modelling results can be triangulated, it is suggested that the LEPs should be required to collect supporting evidence on local impacts as part of their own evaluation activity. There will also be a need to assemble evidence on transport efficiency aspects to support the overall economic evaluation, at least for the more major schemes.

Transport – Urban Sustainable Transport

The proposed approach broadly follows that proposed in relation to roads, involving the construction of matrices of pre and post-investment journey times by public transport between relevant origin/destination (o/d) pairs and their development into an instrumental variable to be used in a fixed effects model to assess the extent to which the investments explain variations in economic performance between locations at varying distance from the projects concerned.

This represents a relatively novel application of the approach and there are questions about whether/how far aspects such as changes in service frequency/quality and improvements in opportunities for walking and cycling can be taken into account. However, there is a potential variation based upon the use of the DfT Access to Town Centres Indices as an alternative instrumental variable if the construction of the o/d matrices proves problematic.

Reflecting both inevitable uncertainties about the extent to which the modelling will be able to identify what may be limited and complex effects on local economic activity and doubts about the likely usefulness of secondary data on important metrics such as rentals and local development activity, it is suggested that LEPs should be required to collect monitoring data on these metrics for both the treatment areas/corridors and potential comparators as part of the local evaluation effort. Again evidence on transport efficiency benefits will be needed to feed into the economic evaluation.

Transport – Rail

An analogous approach to that proposed in relation to other transport schemes based upon estimating the effects on journey time savings by rail between relevant o/d pairs, feeding the findings as an instrument into a fixed effects modelling framework and

assessing the extent to which projects appear to improve the performance of locations in the vicinity of the stations involved relative to those in progressively more distant 'rings' is suggested. This is subject to similar caveats about the novelty of this application of the method, the uncertainty about whether many of the investments involved are likely to generate impacts of sufficient scale to be identifiable with any confidence and the weaknesses of some of the relevant datasets.

Again, it is therefore suggested that the LEPs should be expected to undertake supporting monitoring and evaluation activity through assembling data on changes in rentals and on development activity around the stations involved and potential comparators. Similarly, evidence on transport efficiency aspects will be required to support the economic evaluation.

Skills – Capital projects

The proposed approach is to adopt a matching comparator design, contrasting performance of assisted and non-assisted colleges, taking into account a variety of other attributes including previous receipt of capital support. In the event that this design is not feasible, it is proposed to follow previous BIS studies using a longitudinal panel modelling framework to assess the impacts of the investments on student numbers and other aspects of attainment, at beneficiary colleges. In either case other evidence from BIS sponsored studies based upon a mix of primary research and literature reviews can then be used to assess the likely consequential economic impacts of the identified effects on learner numbers and attainments.

It needs to be noted that the previous studies found it necessary to collect primary data from colleges for the analysis, partly at least in order to address identified weaknesses in some key datasets, particularly the Individual Learner Records. In addition, the passage of time will clearly mean that some of the wider evidence for the impact assessment is likely to need to be 'refreshed' if this is not done as part of other studies for the Department.

Site development – Employment

The projects being funded here appear to have strong similarities to those being funded under the Regional Growth Fund and it clearly makes sense to use the approach which was proposed in the scoping study for the programme which, in turn, is firmly based in the recent academic literature. Essentially this involves assessing gross impacts by using areas within concentric rings of increasing distance from the subject sites as comparators, including within the analysis as controls other potential variables which might contribute to observed performance differences.

Datalinking via the VML can usefully be incorporated to explore the extent to which occupiers are new firms or relocations and, if the latter, whether the relocation involved growth – helping to inform an assessment of how far the identified gross effects involve local displacement. There is an additional issue in this case of how far the Growth Deal investment was essential to the sites' development and possible effects on the timing or scale of the development involved. Realistically this can only be assessed through a case study approach given the impracticability of identifying robust comparators.

Site development – Innovation

The possibility exists of treating these projects in the same way as general employment schemes to assess their spatial impacts and it is considered that this will be worthwhile. However, the more important evaluation question in this case is the extent to which the provision of innovation support has been effective in encouraging the development of successful innovation based businesses. The likely strong element of (self and policy based) selection bias in the beneficiary group rules out any simple comparison group approach and we believe that the most viable strategies will be the use of:

- Firms coming later to the development of the facilities as controls for those coming earlier based on the assumption that the two groups will share similar characteristics; and/or

- Businesses which have sought or are on waiting lists for accommodation as controls based upon similar assumptions.

In either case it will enhance the analysis to incorporate survey work to obtain further details of the business characteristics of the treatment and control groups to improve matching and to provide measures of innovation activity and performance for the two groups which will not typically be available via the VML. Inputs from the LEPs and/or operators of the facilities will be needed to secure contact details for businesses which have sought or are on waiting lists for accommodation and perhaps to help assess their suitability as controls.

Site development – Housing

The importance of planning policies in shaping the pattern of housing development and the time path of price changes, along with the limited extent to which new housing is likely to drive very local economic change and the likely fairly unique character of many housing sites, mean that any modelling based approach to the definition of counterfactuals will lack credibility. In this case we therefore propose that the evaluation should be based upon a combination of:

A set of case studies to explore the extent to which Growth Deal funding was crucial to any housing development on the sites concerned or influenced the scale, form or timing of the development. Where housing formed – or potentially could have formed - a component of a mixed use development, there is clearly an issue of how the availability of Growth Deal funding may have influenced the development mix (though this is most likely to have enabled schemes to incorporate a larger employment component with less housing); and

Aggregation of the housing outputs of various types associated with the Growth Deal schemes.

Business support

The most viable strategy for the national evaluation is likely to be the use of later assists as a comparison group for earlier assists. If this proves problematic because of sample sizes or timing issues, an alternative would be a longitudinal panel (fixed effects) approach, potentially including unsuccessful as well as successful applicants to increase

numbers of observations with dummy variables reflecting the different types of support involved if numbers are too small to allow separate evaluation of different types of scheme. Either approach will benefit if LEPs are able to provide supporting data on applicant characteristics to improve matching or provide potential control variables.

At individual LEP level there may be potential to utilise RDD type methods, depending on the numbers of observations available.

Flood management

The approach proposed takes advantage of existing practice in terms of evaluating the impacts of flood events. Just as footprints of flood damage can be identified, footprints of (additional/modified) flood protection afforded to business and homes can be identified as part of Growth Deal investment programmes. Examining the profile of impact metrics as a result of changes in risk levels will provide a basis for assessment.

The key element of the approach is a capacity to identify pre/post intervention patterns of flood risk and, where necessary, graduated change in level of flood risk, on an annual basis if schemes are implemented over a long-period of time. Flood risk will need to be mapped against administrative boundaries to permit datalinking and account will need to be taken of any previous flood-related investment at or close to the location of intervention. It may be that collection of this information will provide the basis of emerging evaluation techniques at a later stage.

Digital / broadband

It is proposed that the impact of broadband investment be examined through an approach which bears some similarity to recent academic studies in the area. In effect, the approach operates by contrasting the impact of changing coverage and speed across comparator areas and examining whether there is evidence of differential impact as investment is rolled out.

As with flood management, the key element of the approach is a capacity to identify pre/post intervention patterns of broadband coverage and speed on an annual basis which can be mapped against administrative areas to permit datalinking. The construction of coverage/speed maps is an area of uncertainty though academic studies have been able to access quite detailed information with regard to local capacity and the Ofcom fixed broadband dataset provides capacity/speed information down to postcode level. Some attention may need to be paid to the extent of pre-intervention trends but collection of this information will again also provide a basis for use of emerging evaluation techniques at a later stage.

References

Ahlfeldt, G., Koutroumpis, P. and Valletti, T. (2014) Speed 2.0 Evaluating Access to Universal Digital Highways, SERC Discussion Papers, No. 161. Spatial Economics Research Centre, LSE.

Bia, M. and Mattei, A. (2007) Application of the Generalised Propensity Score. Evaluation of public contributions to Piedmont enterprises, Dipartmento di Politiche Pubbliche a Scelte Collettive, POLIS, Università Del Piemonte Orientale.

Criscuolo, C., Martin, R., Overman, H. G. and Van Reenen, J. (2012) The Causal Effects of an Industrial Policy, SERC Discussion Papers, No. 98. Spatial Economics Research Centre, LSE.

DiNardo, J. and Lee, D. S. (2010) Program Evaluation and Research Designs, NBER Working Papers 16016, National Bureau of Economic Research, Inc.

Dubé, J., Legros, D., Thériault, M. and Des Rosiers, F. (2014) A spatial Difference-in-Differences estimator to evaluate the effect of change in public mass transit systems on house prices, Transportation Research Part B: Methodological, Volume 64, June 2014, Pages 24-40.

Faggio, G. (2014) Relocation of Public Sector Workers: Evaluating a Place-based Policy, SERC Discussion Papers, No. 155. Spatial Economics Research Centre, LSE.

Garcia-López, M., Holl, A. and Viladecans-Marsal, E. (2015), Suburbanization and highways in Spain when the Romans and the Bourbons still shape its cities, Journal of Urban Economics, Elsevier, vol. 85(C), pages 52-67.

Gibbons, S., Lyytikainen, T., Overman, H. G. and Sanchis-Guarner, R. (2012), New road infrastructure: the effects on firms, SERC Discussion Papers, No. 117. Spatial Economics Research Centre, LSE.

Gibbons, S., Lyytikäinen, T., Overman, H. G. and Sanchis-Guarner, R. (2012), New Road Infrastructure: the Effects on Firms, SERC Discussion Papers, No. 117. Spatial Economics Research Centre, LSE.

Gibbons, S., McNally, S. and Overman H.G. (2013) Review of government evaluations: A report for the NAO, National Audit Office.

Gibbons, S., Nathan, M. and Overman, H. G. (2014) Evaluating spatial policies, Town Planning Review, 85 (4). 427-432. ISSN 0041-0020

Gibbons, S., Overman, H. G. & Patacchini, E. (2014) Spatial Methods, SERC Discussion Papers, No. 162. Spatial Economics Research Centre, LSE.

Gibbons, S., Overman, H. G. and Nathan, M. (2014) Evaluating spatial policies, SERC Policy Papers, No. 12. Spatial Economics Research Centre, LSE.

Giua, M. (2014) Spatial Discontinuity for the Impact Assessment of the EU Regional Policy. The Case of Italian Objective 1 Regions (197). Università degli Studi Roma Tre, Rome.

Givord, P., Quantin, S., and Trevien, C. (2012) A long-term evaluation of the first generation of the French urban enterprise zones, Documents de Travail de la DESE-Working Papers of the DESE, vol. 1.

Graham, D.J., McCoy, E.J. and Stephens, D.A. (2014) Quantifying Causal Effects of Road Network Capacity Expansions on Traffic Volume and Density via a Mixed Model Propensity Score Estimator, Journal of the Americal Statistical Association, Vol: 109, pp. 1440-1449. ISSN: 0162-1459

Grimes, A. and Young, C. (2013) Spatial Effects of urban rail upgrades, Journal of Transport Geography, Volume 30, June 2013, Pages 1-6.

Holl, A. (2014) Highways and productivity in urban and rural locations, unpublished manuscript.

Hurst, N. B. and West, S. E. (2014) Public transit and urban redevelopment: The effect of light rail transit on land use in Minneapolis, Minnesota, Regional Science and Urban Economics, Volume 46, May 2014, Pages 57-72.

Malgouyres, C., Mayer, T., Mayneris, F. and Py, L. (2012) Évaluation des effets des politiques ciblées sur les territoires: le cas des Zones Franches Urbaines, Rapport remis là Dares, Sciences Po, Paris.

Overman, H. G. (2014) Commentary: making an impact: misreading, misunderstanding and misrepresenting research does nothing to improve the quality of public debate and policy making, Environment and Planning A, 46 (10). pp. 2276-2282, ISSN 0308-518X.

Overman, H. G. and Einio, E. (2012) The Effects of Spatially Targeted Enterprise Initiatives: Evidence from UK LEGI, ERSA conference papers ersa12p164, European Regional Science Association.

Sanchis-Guarner, R. (2013) Driving up wages: the effects of road construction in Great Britain. SERC Discussion Papers, No. 120. Spatial Economics Research Centre, LSE.

Sherman, L. W., Gottfredson, D., MacKenzie, D., Eck, J., Reuter, P. and Bushway, S. (1997) Preventing Crime: What Works, What Doesn't, What's Promising, Report to the U.S. Congress. Washington, D.C.: U.S. Dept. of Justice, 655 pp.

Van der Steeg, M., Vroomen, B. and Cornet, M. (2006) Do innovation vouchers help SMEs to cross the bridge towards science?, The Hague: CPB.

What Works Centre for Local Economic Growth (2014) Evidence Review Business Advice

What Works Centre for Local Economic Growth (2014) Evidence Review Employment Training

Consultees

Professor Henry Overman, London School of Economics

Professor Dan Graham, Imperial College, London

Mark Abram and Chris Gunning, Department for Transport

Zainab Agha and Richard Turl, Department for Communities and Local Government

Sara Davies, BEIS

Simon Dancer, Homes and Communities Agency

Nigel Guy, Marc Eatough, Patrick Bowes and James Flanagan, Leeds City Region

Katrina Hann, David Morris and Richard Waggett, New Economy Manchester

Additional contacts have been made with Defra, DCLG and DfT officials