



Department for
Business, Energy
& Industrial Strategy

BUSINESS SUPPORT EVALUATION FRAMEWORK

January 2019



© Crown copyright 2019

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. To view this licence, visit nationalarchives.gov.uk/doc/open-government-licence/version/3 or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: psi@nationalarchives.gsi.gov.uk.

Where we have identified any third-party copyright information you will need to obtain permission from the copyright holders concerned.

Any enquiries regarding this publication should be sent to us at: enquiries@beis.gov.uk

Contents

Executive Summary	4
Purpose of the Framework	4
Summary of Key Principles	4
Evaluation Planning	7
Logic Models	7
Timelines	8
Evaluation Plan	8
Quality	11
Design	11
Mixed Methods	13
Data Collection	16
Minimum Data Requirements	16
Permissions	17
Analysis	18
Analysis Plans	18
Outcome Measures	18
Economic Impact Measures	19
Reporting & Deliverables	20
Data Storage	22
Feasibility and Proportionality	23
Factors Influencing Proportionality	23
Ethics	25
Governance	26
Guidance & Further Reading	27

Executive Summary

High quality evaluation is important to government, as it enables an understanding of what works, how, and for who, and whether programmes have met their objectives. It helps build an understanding of what the most efficient and cost-effective ways to achieve an outcome are.

Evaluation is the cornerstone of evidence-based policy making, helping policy officials understand impact, and make better decisions on current and future policies. Robust evaluation informs spending reviews, fiscal events, and enables government to be accountable for departmental spend and to weigh up spending decisions between different policies. Evidence on the effectiveness of business support is crucial for informing those decisions.

Purpose of the Framework

Evaluations of Business Support interventions funded by the Department for Business, Energy and Industrial Strategy (BEIS) are expected to adhere to this framework, including current and future initiatives¹. This framework is intended to assist policy makers, analysts, evaluators and delivery bodies to understand what standards are expected and to maximise comparability across different programmes. Programmes vary in nature, therefore it is important to generate evidence which is comparable across multiple ways of addressing the same problem. Evidence generated from these evaluations will help build up a picture of the relative impact and cost-effectiveness of these programmes and their component parts, to inform future policy decisions.

Impact, process and economic evaluation are all recognised as being important in understanding what works, however this framework focuses primarily on impact evaluation. The framework is not intended to be a 'how to guide' for evaluations, as guidance already exists for this purpose (e.g. the Magenta Book)².

As evaluations of BEIS funded business support programmes are delivered – and as the evidence base builds – lessons will be learnt on what high quality evaluation looks like, meaning this guidance will be iterative and updated periodically.

Summary of Key Principles

In order for evidence to be comparable across schemes and of sufficient quality to inform policy making decisions, a number of necessary principles have been identified. These principles are discussed in more detail in the framework, but in summary:

¹ Please note this does not cover innovation policy. There are existing evaluation frameworks for Innovate UK and for catapults, see references and additional guidance.

² For guidance on how to design and run evaluations please see The Magenta Book and the BIS Evaluation Guidance for evaluating interventions with businesses (in the references and additional guidance).

- Evidence needs to be of sufficient quality for all evaluations, in order to provide convincing evidence of what works.
- Evaluations need to clearly map out the anticipated outcomes of the intervention and identify the appropriate outcome measures that will provide a reliable and meaningful assessment of impact (through developing a logic model and/or theory of change).
- The evaluation needs to be designed to be able to detect impact. The design needs to be able to demonstrate that the intervention is causing any change in outcome measures (e.g. productivity) or at least intermediate or proxy measures (e.g. adoption of new management practices, or self-reported improvements in efficiency). Considerations around the anticipated size of the impact needs to be factored into the design phase, to ensure the appropriate methodology is used.
- Evaluations should be accurate in detecting impact, including controlling for any biases. Where biases are uncontrollable, these should be clearly reported on and caveated in evaluation plans and reports. Some techniques (in some circumstances) may overestimate or underestimate the impact of a programme (e.g. not controlling for selection bias).
- Similar metrics need to be collected across different interventions where possible, so that the department is able to compare like with like (see Data Collection for further details on proposed metrics).
- Measurement tools need to be high quality and consistent where possible. For example, the phrasing of a survey question can influence the answer given and may elicit different information, particularly when using a slightly different definition of a concept.
- Evaluations should collect and calculate the same final impact measures (e.g. productivity) to ensure comparability across evaluations. Administrative data can be used to track longer-term outcomes, where it is not feasible to collect these measures directly. Evaluators will need to clearly label what metrics will be reported on from the outset (e.g. defined in the logic model or theory of change), and clearly report how outcome measures are calculated. This will help to ensure that appropriate comparisons can be made.
- Intermediate outcomes, as defined in the logic mode/theory of change, are likely to differ for each intervention³. However, some level of standardisation where possible is desirable.
- Evaluation plans and technical reports need to be detailed and transparent, including clear write-ups of planned and completed evaluation activity, so that informed comparisons can be made between the different interventions. This includes a logic model or theory of change, overarching objectives, independent variables to test for impact and when measures will be collected.
- Capturing and publishing details on the interventions themselves is important, so that intervention designs can be compared and contrasted. This is also important for understanding how the results have been achieved and could be replicated where an intervention is effective.

³ BEIS does recognise that it will not always be feasible for all evaluations to collect the same data in the same format, and that this may not be proportionate for smaller, lighter touch interventions.

- Similarly, it is important to know which businesses the intervention was targeted at, and the characteristics of the businesses receiving the intervention.

In summary, the purpose of the BEIS Business Support Evaluation Framework is to:

- Ensure good quality evaluations are produced, by establishing minimum standards for evaluating business support interventions, and providing transparency on quality standards.
- Ensure comparability across evaluations of business support interventions through standardising quality, data collection, and reporting.
- Allow new ideas to be explored, to identify the most effective policy ideas and to share findings on what works, leading to the wider adoption of more effective policies.

Evaluation Planning

For good quality evaluation to take place, evaluation should be built in during the design phase of a policy or programme, and before implementation. Failure to build in evaluation design early may rule out some of the most robust evaluation designs and additionally may mean that important data is not collected, reducing the quality and credibility of the evaluation.

Sufficient budget, time and staff resources should be made available for good quality evaluations to be conducted. The amount of budget and resource allocated will vary by programme and is dependent on the factors described in the feasibility and proportionality section. The Magenta Book⁴ provides further explanation on resources required and what to consider when making this decision.

Evaluations will ideally be carried out by an independent evaluation partner, rather than the organisation delivering the policy or programme. However, in some circumstances this may not be practical or necessary. Where the same organisation is carrying out the intervention and the evaluation, a description of how independence will be maintained should be included in evaluation plans and reported in the technical write-up. Whether independent or not, evaluators need to have the necessary skills and capacity to carry out the evaluation.

Logic Models

Evaluation plans should state clearly the aims and objectives of a policy or programme and include a logic model or theory of change to underpin the design of the evaluation. A logic model will help to see the intended mechanism for change for each programme or policy, and should also underpin what data needs to be collected and when for each evaluation. Logic models should also help to identify risks or contingency factors, which may mean outcomes are not as envisaged. See a high-level structure of a basic logic model structure below. For more information on theory of changes, see guidance by The Centre of Theory of Change⁵.

Figure 1: High-level structure of a logic model



The key components of a logic model are as follows:

- **Inputs:** Resources used for a programme or intervention (e.g. money and people).

⁴ See the Magenta Book

⁵ See Guidance from Centre of Theory of Change in references and additional guidance.

- **Activity:** What is being delivered e.g. a training programme, events, consultations with businesses.
- **Outputs:** Measurement of what has been delivered e.g. the number of businesses who have completed training, programmes, attended events, received a consultation.
- **Outcomes:** Measurement of any changes for key stakeholders, positive and negative, including economic, environmental, social and personal, (e.g. increased knowledge or capability for managing the business).
- **Impact:** Net change for key stakeholders, including wider economic and social impacts.

Timelines

Evaluators should set out the appropriate times to collect and analyse data on short-term, medium-term, and long-term outcomes and when they would expect to see impacts. This timeline should be presented as part of the logic model and/or theory of change. Timing should be evidence based, using existing literature about similar interventions or, where this does not exist, through consultation or primary research with businesses.

A report by the Enterprise Research Centre (ERC) suggests that evaluations of business support interventions have often looked at too short a time period to see impacts; they recommend monitoring impact for a minimum of 3 years⁶. Three years of good administrative data should be collected as a minimum to understand short to medium-term impacts, and plans should be made to follow up and assess longer-term, ultimate impacts post 3 years.

When thinking about analysing longer term impacts in administrative data, evaluators should consider time lags in administrative data, as well as when impacts would be evident in the business. This could mean conducting the analysis for longer than 3 years post intervention, as administrative data may lag events by up to 2 years.

As a guideline, BEIS would consider the following timescales to be short, medium and long-term, however, the exact timing should be based on the specific intervention.

- The short-term < 6 months
- The medium-term 1 – 3 years
- The longer-term > 3 years

Evaluation Plan

Detailed plans or protocols for evaluations will help to monitor the quality across programmes and interventions. The list below⁷ should be included in a fully developed evaluation plan, or there should be an evaluation protocol which can be developed at a later stage.

⁶ See Drews & Hart, 2015 in the references and additional guidance.

⁷ This has been adapted from the SPIRIT statement checklist which was developed for medical trials (see references and additional guidance).

All BEIS Business Support evaluations should have a detailed plan/protocol which can be reviewed by BEIS and/or independent experts. Creating a plan at the outset of an evaluation enhances the credibility of an evaluation, ensures that key elements of an evaluation are considered from the outset, and provides a record of decisions made which can be particularly useful when there are multiple parties involved in the evaluation or where there are staff changes.

Introduction

- Background and rationale of the policy and the evaluation (including potential users of the findings).
- Primary and secondary objectives.
- Evaluation questions and overarching methodology.

Participants, interventions, and outcomes

- Evaluation Design.
- Study setting (both in terms of geographical location and, if applicable, where an intervention would be administered, such as on business's premises).
- A clear description of the interventions for each group, including how and when they will be administered, with enough detail to allow replication.
- The customer journey for businesses in the evaluation.
- Criteria for modifying or discontinuing interventions (if applicable). For example, if there is significant evidence of harm either to those receiving, or those not receiving, the intervention or if the number of programme applicants is significantly lower than expected.
- Where relevant, strategies to improve adherence or fidelity to an intervention (such as instructions for delivery of the intervention or training), and any monitoring of adherence.
- Primary, secondary, and other outcomes (these should be linked to the logic model/ theory of change).
- Specific measurement variables for the outcomes (e.g. productivity) and how these will be calculated (e.g. turnover/number of employees).
- Sample size – estimated number of businesses needed for the evaluation and how it was determined (including calculations and assumptions). When considering this, policy makers and evaluators should consider what the minimum effect size⁸ they would expect to see and how this relates to the sample size chosen (i.e. is the sample size large enough to detect this effect). Consideration of whether long-term follow up of businesses is important and the extent of attrition in the sample over time.
- Strategies for recruiting businesses.

⁸ Effect size is an objective measure of the size of an observed impact or effect.

Methods for identifying and assigning a treatment and control group

- Method of randomisation for randomised controlled trials.
- Methods of selecting a control group for non-randomised trials e.g. Regression Discontinuity Design (RDD) scores used and cut off, Propensity Score Matching (PSM) or alternative matching technique for matching⁹.

Data collection, management and analysis

- Plans for collecting baselines, intermediate, and outcome data and any other research to be conducted (e.g. qualitative interviews), including who is responsible for data collection. This should include plans for collecting data on unintended effects of the intervention (these will be specific to each programme and should be thought through alongside the logic model/theory of change).
- Timeline for data collection and when data will be available.
- Particular subgroups of interest (for example, micro firms, or a particular sector of interest), which could be driven by the aims of the programme and/or the nature of the intervention.
- Permissions for data use, including permissions for data linking.
- Data storage and data security.
- Plans to promote business retention in the programme/intervention and to complete follow-up, including a list of any outcome data to be collected for businesses who drop out of programmes if applicable.
- A description of interim and final analysis to be undertaken, including statistical methods for analysing primary and secondary outcomes, planned sub-group analyses and how missing data will be handled.

Ethical considerations

- Any ethical concerns (for example, any harm caused to the businesses who receive the intervention, or to a comparison group who do not receive the intervention), and how these will be mitigated. Please see the Innovation Growth Lab (IGL) trials toolkit (pre-trial preparation section) for information on ethical issues specifically in trials¹⁰, and see ethics section within this document.
- Processes for ensuring data confidentiality.
- Processes for obtaining consent from businesses to be part of the evaluation.
- Declaration of interests – any competing interests of evaluators should be declared.

⁹ See Magenta Book supplementary guidance on impact evaluation for further information on RDD and PSM.

¹⁰ See Innovation Growth Lab guidance in references and additional guidance.

Quality

High quality evaluations are well designed and use high quality measurement tools. Expectations for these quality factors are set out below.

Design

Robust impact evaluations should obtain strong evidence of how much any change in business behaviour or performance (if any) can be attributed to a specific policy or programme. Evaluators should consider the strongest designs possible for establishing causality for BEIS Business Support evaluations. This means starting with considering the feasibility of a Maryland Scale 5 evaluation and then moving down the scale where this is not feasible or proportionate (see later section on feasibility and proportionality). Where Maryland Scale scores of 4 or 5 are not feasible, BEIS's minimum expectations for quantitative evaluation designs are in line with the What Works Centre standard of a minimum of Maryland Scale Score 3 (matching with before and after)¹¹. This would include 4 main evaluation designs: Randomised Controlled Trials (RCTs), Regression Discontinuity Design (RDD), Instrumental Variables (IV), and Matching. Other Maryland 3 methods are also acceptable where RCTs or RDD cannot be used, including **Difference in Difference** and **Regression** analysis.

BEIS funded Business Support Evaluations which claim impact must:

- Include a comparison group
- Collect before and after data
- Collect data for both treatment and control group

Table 1 below sets out expectations of which design should be used when. This is a simplified typology, so evaluators should seek guidance from BEIS evaluation experts when the situation is unclear.

Table 1: Selection of an Appropriate Evaluation Design

Factor 1	Factor 2	Expected Evaluation Method
Evaluation design is built in from the beginning (this is expected for all new or recent policies and programmes) and it is possible to withhold the intervention from some businesses (e.g. Business Basics).	Randomisation possible and acceptable	Randomised Controlled Trials (RCT) with before and after data.
	Randomisation not possible or acceptable	Regression Discontinuity Design (RDD) , with before and after data. Where randomisation is not possible or acceptable, then allocation could be based on a score, allowing the possibility of

¹¹ See What Works Centre for Local and Economic Growth guidance on the Maryland Scale in references and additional guidance.

Factor 1	Factor 2	Expected Evaluation Method
		<p>using RDD. However, this design relies on more assumptions than an RCT and it also requires a lot of businesses to fall around the cut-off in order for it to be feasible. Therefore, randomisation is preferable where possible.</p>
<p>Where it is not possible to withhold interventions from some businesses or in the case of existing programmes or policies where allocation has already occurred (and was not based on a score or randomisation). But where there is sufficient data to track outcomes using administrative data, (e.g. Growth Hubs).</p>	<p>Presence of a convincing instrumental variable.</p> <p>No appropriate instrumental variable</p>	<p>Instrumental Variable (IV), with before and after data. Where there is a suitable instrumental variable which correlates with the allocation but not the outcome measure, IV can be applied. This technique can be applied retrospectively.</p> <p>Matching, with before and after data. Matching is less able to demonstrate causality than RDD or RCTs, as there is the problem of selection bias. However, it does at least control for observable factors. Given the problem of selection bias, evaluators should consider whether there are observable factors which may control for selection bias.</p> <p>Other Maryland 3 methods are also acceptable where RCTs or RDD cannot be used, including Difference in Difference and Regression analysis.</p>

Selection bias may be a significant confounding factor in the performance of businesses applying to business support schemes, as there are likely to be factors which are difficult to observe (e.g. ambition, leadership and management, and networks). Given this, where matching is used, this limitation needs to be made clear, and evaluators should consider what they can do to minimise the impact of selection bias.

It may be that some elements of a programme or policy can be evaluated more robustly than the programme or policy as a whole. This can still generate valuable evidence and should be considered alongside the more holistic evaluation in these cases. Variation in treatment can

also be considered where different types of support (for example light touch compared to intensive) are compared. This may be more practical and more acceptable than using business-as-usual as a control. This particularly applies to randomised controlled trials.

If there is a clear rationale as to why it is not possible to meet this minimum standard of matching, other evaluation methods can be considered (e.g. before and after designs) and should be discussed with a BEIS evaluation expert. As these types of designs do not control for macro factors or selection bias, these should only be considered in situations where there is no appropriate or feasible comparison group.

Qualitative Evaluation

Although this framework focuses on quantitative impact evaluation, the quality of how a qualitative evaluation is carried out is just as important. Qualitative evaluation should be used in addition to quantitative evaluation where appropriate. Please refer to the Magenta Book¹² and supplementary guidance on qualitative evaluation for further information on conducting qualitative evaluations and the standards that should be applied to qualitative evaluation methods.

While this framework sets a minimum standard for impact evaluations, this requirement should not undermine the importance of other monitoring and evaluation activities. Rigorous monitoring is always important and should still be conducted even in circumstances where evaluation is not planned. Similarly, economic and process evaluation is important in understanding the costs and mechanisms which underpin any impact (or absence of impact), which should be considered through the development of a logic model (see above).

Other Considerations

Evaluations need to consider how to address the following and should justify any adjustments made based on existing literature on business support, and following guidance laid out in the Green Book¹³:

- Deadweight
- Displacement
- Substitution
- Spill-overs
- Leakage and differences between national and regional impacts

Mixed Methods

For anything other than very light touch interventions, a mixed methods approach to evaluation should be used to improve the quality of the evaluation. Qualitative research adds more depth to evaluation and can help in answering the how and why questions, which may not be answered through analysis of the quantitative outcome measures.

¹²See the Magenta Book, in the references and additional guidance section.

¹³See the Green Book, in the references and additional guidance section.

Qualitative evidence will help explain why an intervention has or has not worked, identify any unintended consequences, identify barriers and enablers to participation and facilitate a deeper understanding of scheme impacts. For example, understanding whether impact varies between individuals or sub-groups, and gaining a better understanding of the programme's contribution to impact (e.g. collecting data on other support business are receiving).

Qualitative research, management information, and surveys can also shed light on the fidelity of an intervention, and whether it was delivered as planned. How an intervention has been implemented may be essential as to whether an intervention works or not. What support has been delivered in practice, as well as what should have been delivered in theory, should be recorded.

In addition, feedback and data at an early stage can help to improve the implementation of an intervention. It may also be worth considering carrying out qualitative research before policy design is finalised, in order to help establish the best way to measure impacts and what the most important things to measure are in terms of delivery and impact. For many interventions, it will also be necessary to conduct primary research such as quantitative or qualitative surveys, in order to collect interim measures. All research carried out should be in line with the Government Social Research (GSR) Code of Practice¹⁴.

Considerations for Survey Quality

For surveys, the evaluator needs to ensure high quality survey research is being collected, including considering the following factors:

- **Coverage of the treatment and control group** - if taking a sample of the treatment and control group, this should be a representative sample using a random probability-based sampling method where feasible. Depending on the size of programme, a census of beneficiaries may be appropriate. Where random sampling is not appropriate, an alternative sampling method should be agreed (e.g. quota sampling).
- **Coverage of subgroups of interest** - such as firms of different sizes and sectors, and ownership models.
- **Response rates** – efforts should be made to maximise survey response rates in order to minimise non-response bias¹⁵.
- **Question design** – pre-tested, standardised questions should be used where possible in order to ensure quality, and to maximise comparability across different interventions¹⁶.

Considerations for Qualitative Research

Qualitative research should be guided by the principles for good quality qualitative research. More generally, including a defensible design, rigorous, systematic and transparent data collection, analysis and interpretation, and credibility of claims made from the research.

¹⁴ See GSR Code of Practice in the references and additional guidance.

¹⁵ Non-response bias can also be conducted to understand any limitations in the data (e.g. if certain sub-groups are less likely to have taken part).

¹⁶ See the UK Data Service for a database of existing questions, and the ONS Annual Business Survey for questions used to calculate productivity in the references and additional guidance

For more information on Process Evaluation see The Magenta Book and supplementary guidance, which includes an overview of qualitative methods¹⁷.

¹⁷ For more information see the GSR publication 'Quality in Qualitative Evaluation: A framework for assessing research evidence'.

Data Collection

Minimum Data Requirements

In order for comparisons to be made between different policy interventions, BEIS requires similar data to be collected across these policies or programmes. The list below details the minimum data requirements in order to facilitate this.

Identification of Programme Participants

Where possible, this should include unique business identification numbers to allow the data to be matched with administrative data sources (e.g. HMRC) for the purposes of longer-term economic analysis. This could be one of the following:

- Companies House Registration Number (this should be collected as a preference and should be verified using the Companies House website).
- VAT Number.
- PAYE Number.
- Unique Tax Reference Number in instances when the business is not yet VAT or PAYE registered.

Identification numbers for the recipients of BEIS business support policies are particularly important, as this allows the department to track these businesses over time using administrative data. It may not always be possible or practical to collect unique business identifiers (see feasibility of data collection below). In this instance, please discuss with BEIS evaluation experts the best course of action. Identification includes the following:

- Business Name.
- Trading name (if different).
- Address.
- Postcode.
- Business contact details (name, surname, landline number, mobile number, email address).
- Age of business (or trading since).
- SIC code.
- Number of employees.

- Baseline measures for turnover, Gross Value Added (GVA) and productivity or Total Productivity Factor (TPF)¹⁸.

Impact Measures

- Inputs – record of programme costs and resources.
- Activities – intervention delivered e.g. training.
- Outputs – measures of what has been delivered, for example the number of businesses receiving support and what they received.
- Outcomes – measure of change for key stakeholders, positive and negative, including economic, environmental, social and personal, (e.g. increased knowledge or capability for managing the business).
- Impacts - Net change for key stakeholders, including short-term, intermediate and long-term measures of impact.

Permissions

For all evaluations, it is important to obtain required permissions from those being evaluated. General Data Protection Regulation (GDPR) requirements¹⁹ need to be met, and where possible, informed consent is preferred. Linking data may be carried out by the department for evaluation and research purposes; this should also be made clear to programme applicants.

¹⁸TFP is a measure of how efficiently all inputs are converted into output produced. It considers the difference in output produced that traditional inputs such as labour (workforce) and capital (machinery, tools) fail to explain. TFP is a more robust measure of productivity, so should be collected where feasible, but BEIS recognises that this isn't always possible.

¹⁹ See the EU GDPR guidance in the references and additional guidance.

Analysis

Analysis Plans

Analysis plans should be developed at the start, prior to carrying out the evaluation. This can be incorporated into the evaluation plan or can be separate (see evaluation plan section for more detail).

All analysis should be quality assured by someone independent (i.e. someone who was not involved in carrying out the original analysis). Quality assurance processes should be clearly outlined in evaluation plans.

Outcome Measures

All evaluations should report outcomes in line with the theory of change or logic model. These should include short, medium, and long-term measures so that the project manager can identify at an early stage, whether or not the intervention is working. All evaluations should collect the following measures:

- GVA
- GVA/per hour worked or GVA/worker
- Proxy measures such as Turnover/Headcount should be considered where GVA/per hour worked or GVA/worked is not possible to calculate. It should be clear that this is a proxy measure for productivity and termed turnover/head or similar rather than labelled productivity.
- Turnover
- Employment

Where a proxy measure has been selected over more robust measures, evaluators should describe any limitations of the selected measures. It would be desirable – where possible – to collect Total Factor Productivity (TFP), but its recognised that this is harder to measure and not always feasible.

In the short and medium-term, evaluation outcome measures based on self-reporting/survey evidence should be reported, because using administrative data will not provide timely evidence of impact.

In the longer-term, administrative data should be used where appropriate in order to calculate long-term impacts if more direct measures are not available. For less expensive or more light touch interventions, measures of increased efficiency or uptake of new technologies or practices may be sufficient and may be the most appropriate measures.

Where interventions are high intensity and/or where they are applied to large numbers of businesses, impacts are more likely to be detectable, so long-term outcomes should be assessed through administrative data. Low intensity interventions applied to small numbers of businesses are less likely to produce observable impacts in the administrative data. In order to

assess whether it is likely that the effect will be observable in the administrative data, evaluators should consider the expected effect size, along with the number of businesses receiving the intervention. If the expected effect size is not known, this can be determined through looking at the literature or based on working back from the cost of intervention to determine what would be a success considering the costs.

Historic data can be used to calculate the volatility in outcome variations. Combining this information with the effect size, and the sample size, will give an indication of whether the impacts of the intervention are detectable from the administrative data. Volatility of the administrative data is something that BEIS can assist on, so seek advice from BEIS on this when designing an evaluation. The Magenta Book provides further guidance on this, a section on 'the signal and the noise' provides guidelines on the ratios between effect size and volatility, and the number of data points required to see an effect. This is an important consideration for business support programmes, as the variance in turnover and employment figures is high, in particular for turnover. This can be controlled for, to some extent, by using data from previous years where available.

In most cases, the longer-term impact assessment using administrative data is likely to be conducted by BEIS or commissioned after the initial impact evaluation. Therefore, in the design stages, it is important to ensure the evaluation data can be linked to administrative data by collecting the relevant information as outlined above (e.g. companies house reference number). For some interventions, this may be problematic (see feasibility and proportionality) and in these cases please discuss with BEIS evaluation experts the best approach.

Economic Impact Measures

All evaluations should include Benefit-Cost Ratios (BCRs) or, where these are not possible, Cost-Effectiveness calculations. These should be based on (in order of preference):

- Administrative data (including costs of delivering the programme).
- Self-report data which can be used to generate estimates of GVA.
- Other existing evidence of return on investment. The use of this evidence needs to be fully explained and justified so that the strength of these claims can be assessed.

Discounting factors should be applied, and economic costs will need to be considered. Any assumptions used need to be clear and transparent. For more information on BCRs, see the Treasury Green Book and consult with BEIS appraisal experts.

Reporting & Deliverables

To maximise learning, and to ensure findings are used to inform policy making, all evaluation findings should be published and disseminated in a timely and quality manner to key policy makers and practitioners. BEIS expects the following as standard for reporting and deliverables.

Interim findings

- Balanced summary of short-term and/or medium-term findings including both qualitative and quantitative findings where relevant.

Main findings

- Balanced summary and interpretation of all findings including both qualitative and quantitative findings where relevant at the end of the agreed evaluation period.

Policy Summary

- This should be a succinct, balanced summary of findings from the evaluation with enough detail on the programme and methodology. This should be an accessible overview, without the reader having to consult further documentation.

Technical Reports

- Technical reports should be of a standard level of detail that the research could be replicated by someone not involved in the project, including²⁰:
 - Background and rationale.
 - Objectives of the policy or programme (including the logic model).
 - Detailed description of methodology, including evaluation questions and design and any important changes made to the design following the evaluation plan or protocol.
 - Details on how any qualitative research was conducted, if not included as part of the main evaluation methodology.
 - Methods for identification and assignment of the treatment and control group.
 - Detailed description of the intervention.
 - Eligibility criteria for participants.
 - Baseline demographic information for businesses receiving the intervention (e.g. baseline turnover, number of employees, region).
 - How the sample size was determined and whether this was achieved.

²⁰ This list draws from existing guidance including CONSORT statement, see references and supplementary guidance.

- Study settings (e.g. how and where was data collected).
- A description of whether the intervention was delivered as intended.
- Measures of compliance to treatment (e.g. proportion of businesses offered treatment who accepted treatment, and the proportion who completed treatment where applicable).
- Outcomes – the pre-specified primary and secondary outcome measures and any changes to these measures which occurred after the evaluation plan or protocol.
- Detailed description of statistical analysis conducted, including any tests for the suitability of analytical techniques. This should incorporate analysis of primary and secondary outcomes, any further analysis such as sub-group analysis, and analysis of unintended consequences.
- Syntax used to calculate composite measures and to clean the data.
- Graphs and charts of the data and data tables for surveys (summarising the findings for each question, this can be presented in excel)²¹.
- Research materials such as survey instrument, interview schedule for depth interviews, discussion guides for focus groups.
- Quality measures such as survey response rates, matching rates to administrative data.
- Links to related documents such as the evaluation plan or protocol, policy reports and interim findings (if unpublished, they can be included in the technical report).
- Limitations of the evaluation including sources of potential bias, imprecision, and multiplicity of analyses.
- An assessment of generalisability of the results.
- Details of any ethical considerations.

All reports need to be of a publishable standard, should use the BEIS template where appropriate and should meet accessibility guidelines. See Chapter 10 of the Magenta Book for further guidance on reporting.

²¹ See guidance from the Office for National Statistics for further guidance, in the references and additional guidance.

Data Storage

All data should be handled in line with relevant regulation and codes of practice, including adhering to the Data Protection Act, General Data Protection Regulations (GDPR) and the Market Research Society's (MRS) guidelines.

BEIS should be provided with all the data relating to every evaluation including the raw data set, cleaned data set, syntax or code (including for data cleaning, creating compound variables, and analysis) and extracts from the data set which have been used for analysis by evaluators. Information on how the data will be used should be available to the participant at the time of data collection (including how BEIS will use the data in the future). Evidence of the permissions sought, and information given to programme participants about data should also be given to BEIS with the data set. This will enable BEIS to conduct further analysis and longer term follow up on evaluations of policies and programmes.

Secure data transfer will be required to facilitate this in order to protect the identify of programme participants, and to meet regulatory requirements. The method for transferring data should be agreed with BEIS before data is transferred.

Feasibility and Proportionality

Overall, the department's aim is that evaluations of business support programmes should be evaluated to as high a standard as possible. However, it is also necessary to consider proportionality. For very low-cost interventions which are likely to have small impacts, it may not be value for money to spend a large amount on evaluating the impact. There are also circumstances where data collection may undermine a light touch intervention by causing too much friction in the customer journey. In these circumstances, a pragmatic approach should be taken. Consideration should be given to what administrative and monitoring data is being/can be collected, and whether there are inexpensive methods of data collection which could be employed, such as surveying a sample of beneficiaries. In these circumstances, this should be discussed with BEIS experts to determine the best approach. Considerations for proportionality and feasibility are discussed below.

Factors Influencing Proportionality

Cost

Expensive interventions should be high priority for evaluation from an accountability perspective - in order to ensure the department is making the best use of public resources. The cost of the evaluation should be proportionate to the cost of the intervention. As discussed above, longer-term impacts should be tracked for more expensive interventions.

The cost of the evaluation is also an important factor. Some relatively inexpensive interventions may be easy and cheap to evaluate, so this should also be factored into decisions to evaluate and which design to use. For example, using a variety of different messages to encourage people to use existing tools or sources of intervention is cheap to implement but it is also relatively cheap to trial the different messages.

Intensity and Expected Impact

Lighter touch interventions may be likely to have a smaller impact than more intensive interventions, therefore it may not be realistic to expect long-term effects on productivity to be measurable, except in cases where take up/numbers are very high. This is likely to be closely related to intervention cost, i.e. lighter touch interventions are also likely to be cheaper.

Novelty and Learning Potential

Another criterion for the importance of evaluating a programme is novelty. Novel interventions (or interventions where there is relatively little evidence on effectiveness) should be evaluated as rigorously as possible in order to add to the evidence base on what works.

Similarly, it would be worthwhile investing heavily in one small programme evaluation if results will be relevant to lots of other spending (current or planned).

Pilots

All pilots should be evaluated rigorously – with sufficient resources invested in the evaluation – on the assumption that the pilot is going to inform large future investment or financial cuts.

Controversial or Risky Interventions

Where something is controversial, or where there is potential for harm, the intervention should be evaluated rigorously to minimise this risk.

Evaluation and Data Collection Feasibility

There are many factors which may make an evaluation technique feasible or not. For example, if it is not possible to withhold an intervention from any of the business population (e.g. where something has already been rolled out nationally) then it would not be possible to run a randomised controlled trial. See section above on design. Evaluators will already be familiar with the conditions for using different evaluation methods, however more information can be found in the Magenta Book.

Feasibility of data collection also needs to be considered. For light touch interventions (e.g. an online diagnostic), asking for identifiers (e.g. unique business ID) may interfere with the customer journey and deter businesses from completing the online diagnostic. In these circumstances, it would be ideal to test this to see whether this is the case. If it is deterring businesses from engaging with an intervention, then it may be reasonable not to collect this information. For more intensive interventions, this is less likely to be a problem and, as stated above, more expensive interventions are a priority for evaluation.

Ethics

All research by, or on behalf of BEIS (including evaluations), should comply with appropriate ethical guidelines. For more information on ethical standards see the following guidance in the references and additional guidance section:

- Government Social Research (GSR) Ethical Guidance
- The Economic Social Research Council (ESRC) Research Ethics Framework
- The Market Research Society Code of Conduct

Governance

All evaluation plans should include a governance structure, including clear lines of responsibility and details of how and when key stakeholders will be engaged throughout the evaluation. There should be clear lines of communication with BEIS to be agreed for each project.

Guidance & Further Reading

This framework is not designed to be a 'how to guide for evaluation', as plenty of resources exist to help with evaluation design. Please see below references for the guidance mentioned in this framework, and additional relevant sources of advice and information.

Reference and Description	Link
<p>The Abdul Latif Jameel Poverty Action Lab (J-PAL)</p> <p>Guidance on rules of thumb for determining sample size and power</p>	<p>https://www.povertyactionlab.org/sites/default/files/resources/2018.03.21-Rules-of-Thumb-for-Sample-Size-and-Power.pdf</p>
<p>Centre for Theory of Change</p> <p>Theory of Change benefits</p>	<p>http://www.theoryofchange.org/what-is-theory-of-change/toc-background/toc-benefits/</p>
<p>CONSORT</p> <p>Recommendations for reporting on RCTs</p>	<p>http://www.consort-statement.org/</p>
<p>Drews, C., & Hart, M. (2015).</p> <p>Feasibility Study – Exploring the Long-Term Impact of Business Support Services. Research Paper.</p>	<p>https://www.enterpriseresearch.ac.uk/wp-content/uploads/2015/04/ERC-Research-Paper-LT-Impact.-Research-PaperNo29.-Final-02APR15.pdf</p>
<p>Enterprise Research Centre</p>	<p>https://www.enterpriseresearch.ac.uk/</p>
<p>European Union General Data Protection Regulation</p> <p>Guidance and further information</p>	<p>https://eugdpr.org/</p>
<p>Economic Social Research Council (ESRC)</p> <p>Research Ethics Framework</p>	<p>https://esrc.ukri.org/funding/guidance-for-applicants/research-ethics/</p>
<p>Government Evaluation and Appraisal Guidance</p> <p>The Magenta Book & Supplementary Guidance</p>	<p>https://www.gov.uk/government/publications/the-magenta-book</p>

<p>The Green Book: Appraisal and Evaluation in Central Government</p> <p>The former Department for Business Innovation and Skills (BIS) Evaluation Guidance (2011)</p> <p>Cabinet Office: 'Quality in Qualitative Evaluation'</p>	<p>https://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-government</p> <p>https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/212318/11-1085-guidance-evaluating-interventions-on-business.pdf</p> <p>https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/498321/Quality-in-qualitative-evaluation_tcm6-38739.pdf</p>
<p>Government Social Research</p> <p>GSR Code of Practice</p>	<p>https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/732407/The_GSR_Code_-_Products.pdf</p>
<p>Government Social Research</p> <p>GSR Ethical Guidance</p>	<p>https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/515296/ethics_guidance_tcm6-5782.pdf</p>
<p>Innovation Growth Lab (IGL)</p> <p>Guide and Online experimental toolkit on how to design and run policy experiments in the area of innovation, entrepreneurship and business growth</p> <p>Example of existing trials in the database and list of supported trials</p>	<p>https://www.innovationgrowthlab.org/guide-randomised-controlled-trials</p> <p>http://toolkit.innovationgrowthlab.org/home</p> <p>https://www.innovationgrowthlab.org/igl-database</p> <p>https://www.innovationgrowthlab.org/our-projects</p>
<p>Innovate UK</p> <p>Innovate UK's Evaluation Framework, including an explanation of why to evaluate and covers more detail on different types of evaluation.</p>	<p>https://www.gov.uk/government/publications/evaluation-framework</p> <p>https://www.gov.uk/government/publications/capult-programme-evaluation-framework</p>
<p>The Market Research Society (MRS)</p> <p>Code of Conduct (including ethics guidance)</p>	<p>https://www.mrs.org.uk/standards/code_of_conduct</p>
<p>Office for National Statistics (ONS)</p> <p>Annual Business Survey (ABS)</p>	<p>https://www.ons.gov.uk/businessindustryandtrade/business/businessservices/methodologies/a</p>

	nnualbusinesssurveyabs#example-questionnaires
<p>Office for National Statistics (ONS)</p> <p>Guidance on designing effective charts and tables</p>	https://gss.civilservice.gov.uk/wp-content/uploads/2017/01/GUIDANCE-document-Effective-charts-and-tables-in-official-statistics-Version-2.0-Jan2017.pdf
<p>SPIRIT</p> <p>SPIRIT STATEMENT.</p>	http://www.spirit-statement.org/
<p>UK Data Service</p> <p>Database of existing survey questions</p>	https://discover.ukdataservice.ac.uk/
<p>What Works Centre for Local and Economic Growth</p> <p>The Scientific Maryland scale</p>	http://www.whatworksgrowth.org/resources/the-scientific-maryland-scale/

This publication is available from: www.gov.uk/government/publications/business-support-evaluation-framework

If you need a version of this document in a more accessible format, please email enquiries@beis.gov.uk. Please tell us what format you need. It will help us if you say what assistive technology you use.