NHS England

NHS England Test Beds Programme:

Evaluation learning from Wave 1

Contents


Section 01

Introduction

The Test Beds Programme: harnessing technology, transforming care

The Test Beds Programme is a cross-government venture funded by NHS England, the Department of Health and Social Care and the Office for Life Sciences. It is a pioneering project which combines new digital technologies and models of care in real world clinical settings. The programme generates evidence to inform decisions about where and how the uptake of digital innovations at scale and pace across the health and care system is most likely to improve outcomes for patients and service users at a similar or lower cost. The breadth of digital technologies and changes to patient pathways implemented by Test Beds has been vast, but generally they aim to support self-management of long-term conditions by:

- Empowering patients and carers to take a more active role in management of their own condition.
- Facilitating early diagnosis.
- Preventing unnecessary hospital admissions.
- Bringing care closer to home.
- Supporting frontline workers to deliver care more effectively and efficiently.

Another key feature of the Test Beds Programme is collaboration. Each Test Bed brings together partners from across the NHS, academia, industry, patient groups and charities, who work together to improve patient outcomes.

This handbook brings together the wealth of learning generated from Wave 1 of the programme and includes recommendations and reflections on how best to evaluate these complex interventions. It forms part of a suite of Test Beds legacy and learning publications that seek to share learning from the programme with both future Test Beds and the wider health and care system.

The importance of evaluation to the Test Beds Programme

Evaluation is fundamental to the Test Beds Programme. It provides robust evidence of what worked well and also areas for improvement, in terms of design and delivery of the interventions (process evaluation). It also measures what outcomes were achieved, for whom, and why (impact evaluation) as well as considering the extent to which they offered value for money (economic evaluation). Each Test Bed project has been evaluated by a local evaluation team, generating evidence on the process, impact and economic evaluations. These evaluations have in turn been advised and synthesised by a team of national evaluation partners from Frontier Economics and NatCen Social Research.

The evaluation data produced by Test Beds contributes to a national evidence base regarding the impact of using technological innovations as part of the design of new models of care. The evaluation findings from the programme will be translated into pragmatic learning that commissioners of similar interventions can use to increase the likelihood that these interventions will prove effective and deliver value for money. Most importantly, the programme seeks to support the spread and scale of successful combinations of digital technologies and service change, so that others can benefit from improved outcomes, at the same or less cost to the health and care system.

About this handbook

This evaluation handbook is designed to support a future wave of Test Bed programme directors, healthcare programme commissioners and other Test Bed project staff to:

- Understand common evaluation challenges faced by the initial Wave 1 Test Beds and benefit from their learning.
- Plan similar evaluation activity and access tools and guidance that can support their work.

It is designed to be a reference document for Test Bed leadership teams who will ultimately act as ‘intelligent customers’ of evaluation. Academic and other evaluators are also likely to find it helpful due to the insights and learning identified (though they will inevitably be relying on much more detailed evaluation guidance documents, many of which are signposted in this document). Given the aims and target audience of this document, it is structured around providing the basic information, knowledge and advice to Test Bed leadership teams, informed by learning from Wave 1.

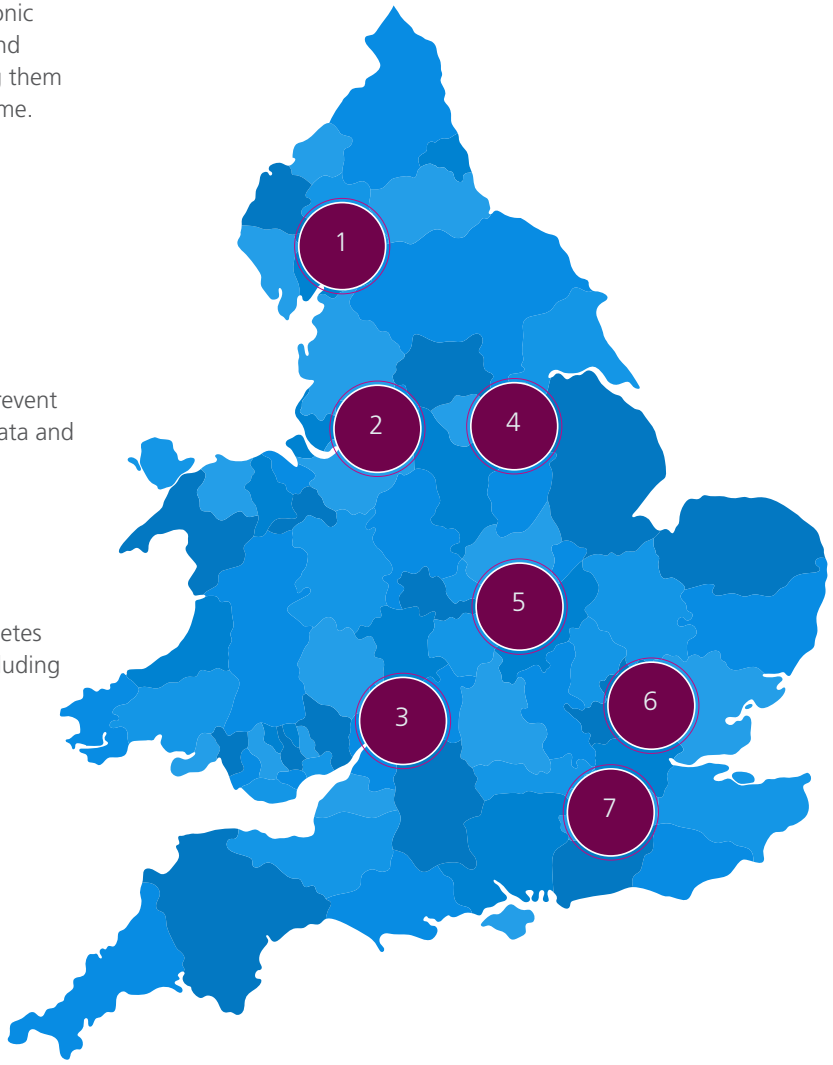
This handbook is therefore structured in four main parts:

- **Section 2:** preparing to commission an evaluation. This draws on Wave 1 to identify the key activities that would help set the evaluation up so that it has clear objectives, scope, stakeholder engagement, governance and resources.

- **Section 3:** commissioning an effective evaluation. Recognising that Test Bed leadership teams are likely to need to commission evaluation either from external parties (or perhaps from an evaluation team within the health and care system), this section offers insights on the processes to procure and select an appropriate evaluation team.
- **Section 4:** being an intelligent customer. This draws on the experience of local evaluators and the National Evaluation Partner to provide a foundation of knowledge for Test Bed leadership teams such that they can test, challenge and understand the evaluation evidence appropriately.
- **Section 5:** learning and dissemination. Evaluation evidence will only be valuable if it is clear, influential and relevant to real-world decisions. This section offers ways in which learning can be actively translated to inform future decisions including the adoption and spread of successful interventions by other NHS Trusts so that more patients can benefit.

The Test Bed sites and the evaluation methodologies used

The map below sets out the evaluation methodologies used at each Test Bed site.



1 **Lancashire and Cumbria Innovation Alliance**

Improving support for those over 55 with Chronic Obstructive Pulmonary Disease, heart failure and dementia. Integrating technologies and linking them to new care models supporting self-care at home.

A D H K L M O Q S T U W

2 **Heywood, Middleton and Rochdale/Long Term Conditions Early Intervention Programme**

Improving the ability to detect, manage and prevent long-term ill health by using pseudonymised data and telecare technology.

A C H L T V

3 **Diabetes Digital Coach**

Facilitating improved self-management of diabetes through the use of integrated digital tools, including the use of Internet of Things technology.

T V

4 **Perfect Patient Pathway**

Improving pathways for asthma, diabetes, falls and frailty by increasing access to technology and facilitating information sharing.

A B F H J K L M

5 **RAIDPlus**

Developing a demand and capacity tool that shows patient flow in real-time and a predictive algorithm to identify when people are going to experience a mental health crisis.

A D F G H I S T V

6 **Care City**

Testing a combination of digital devices and software alongside new approaches to service delivery and patient participation. Example conditions include dementia and falls and frailty related injury.

A F H S T V

7 **Technology Integrated Health Management**

Providing people with dementia and their carers with: wearables, monitors and other devices which will combine into an 'Internet of Things' to monitor their health at home.

A E H J K M N O P R S T V W

Economic / impact evaluation methods

A

Descriptive statistics

This provides insights from collected data by describing the basic features of the sample of patients or service users and the outcome measures for which data is available. This is presented as a narrative as well as by visualising the data.

B

Before versus after

A comparison of outcomes immediately before an intervention with outcomes after the policy has been introduced. In effect, the outcome observed before the policy intervention acts as the counterfactual. It is implicitly assumed that outcomes would remain constant in the absence of the intervention. Any subsequent changes in the outcome are attributed entirely to the intervention.

C

Regression difference in differences (DiD)

This approach follows a similar logic to the simple DiD method, but also accounts for changes that are not as a result of the intervention in the treatment and control groups (i.e. 'controls'). It assumes that in the absence of the intervention, outcomes in the treatment group would follow the trend in the control group, adjusting for changes in either group over the intervention period.

D

Propensity score matching (matched cohort analysis)

This approach aims to ensure that the control group is as similar as possible to the treatment group by using data on the characteristics of patients and matching treatment patients to control patients on the basis of the likelihood that the individual will participate in the intervention given their observable characteristics, (age, gender etc.).

E

Randomised control trial

Patients or users who all meet the eligibility criteria for an intervention are randomly assigned to either a control group or treatment group. As long as the randomisation is effective, the treatment and control groups will be identical, other than the receipt of the intervention. Comparing changes in outcomes between the groups therefore gives a measure of impact.

F

Simulated modelling

A model is developed, using published literature to populate assumptions on key drivers of outcomes, (uptake, behavioural responses, adherence, likelihood of successful treatment etc.) It simulates the choices that can be made at each stage of the patient pathway and the outcomes under each choice.

G

Statistical process control

A tool which supports ongoing monitoring of process or outcome measures during a pilot. It uses simple statistical rules to identify when changes have occurred, e.g. significant improvement or deterioration in outcomes. It can be helpful to provide 'real time' results for formative evaluation rather than waiting until summative evaluation is possible before observing results.

Outcome measures

H

Secondary care activity

Administrative data collected by trusts logging the amount and nature of patient activity over a period of time. For example, measures might include A&E attendances, hospital admissions, and patient length of stay.

I

NHS operational performance

Operational data collected by trusts which measures the performance and efficiency of services. For example, measures might include hospital waiting times, utilisation of beds, or frequency of delayed discharges.

J

Quality of life

This is measured using a number of instruments such as DEMOL (measures quality of life changes for those with dementia); DEMQOL Proxy (measures changes in quality of life for those with dementia, completed by the carer); ReQoL-10 (patient-reported outcome measure for people with mental health conditions); ICECAP-A/0 (patient-reported outcome measure).

K

Health-Related Quality of Life

The EQ-5D5L (EuroQoL) (Dolan et al., 1995) is a simple and well-validated measure that can be completed in less than 10 minutes. It is able to record changes in overarching health-related quality of life (an overall score is provided) as well as in the five areas of mobility, self-care, usual activities, pain/discomfort and anxiety/depression.

L

Psychological wellbeing

This is measured using the Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS). This is used to enable the monitoring of mental wellbeing in the general population.

M

Patient activation

This is measured using Patient Activation Measure (PAM)13. This is a 13-item scale that assesses patients or users' knowledge, skills and confidence in self-managing their long-term condition (Hibbard et al, 2005). The PAM segments individuals into one of four progressively higher levels of activation and successful treatment etc. It simulates the choices that can be made at each stage of the patient pathway and the outcomes under each choice.

N

Activities of Daily Living

This is measured using the Bristol Activities of Daily Living Scale. It is a short, carer-rated measure covering 20 activities of daily living, both basic and instrumental. Items are rated on a 4-point scale (from totally dependent to totally independent, with an additional 'not applicable' option) over the previous two weeks.

O

Patient reported service use

This is measured using the Client Services Receipt Inventory (CSRI) (Beecham and Knapp, 1992). Participants are asked to indicate if and how often they have used primary, secondary and tertiary health services, social care or third sector services and informal care over the previous three months, allowing measurement of overall changes in service use.

P

Carer-related outcomes

This is measured using the Zarit Care Burden Scale. This is the most widely used measure for carer burden in carer research (Moniz-Cook, 2008). It is thought to be useful for identifying carer burden and predicting carer collapse (Gort et al., 2007). The revised version comprises 22 items. Each item is a statement which the carer endorses using a 5-point scale from 0 (never) to 4 (always).

Q

Loneliness

This is measured using the De Jong Gierveld Loneliness Scale which is a commonly used scale translated and validated in several European countries. The Jong Gierveld is an 11 item scale with three response categories, summed to provide a single score ranging from 0-11 (11 being the highest score or highest level of loneliness).

R

Technology acceptance

This is measured using the Technology Acceptance Model (TAM); and Global Attitude Towards Technology (GATT) measure. The TAM (Davis, 1989) is used to predict technology acceptance that has been used successfully by those with dementia (Chen and Chan, 2011). The GATT is a single-item straight forward question that measures generic feelings of discomfort and fear towards new technology using a multiple-choice format.

S

Adherence/compliance

This is measured directly from the app or online tool as it records the number of times the service user has logged in or has utilised the material.

Process evaluation methods

T

Focus groups

Small group discussions among programme delivery staff or recipients which focus on their thoughts and opinions regarding their experiences with the intervention.

U

Interviews

Dialogue based on open questions and topics which include in-depth exploration of how processes work in the implementation across stakeholder groups. (I.e. who, what, when, where, why, how.)

V

Observations

An open data collecting tool which involves the evaluator taking detailed field notes about the implementation aspect or an event.

W

Document review

An assessment and analysis of documents linked to the design and implementation of the intervention.

X

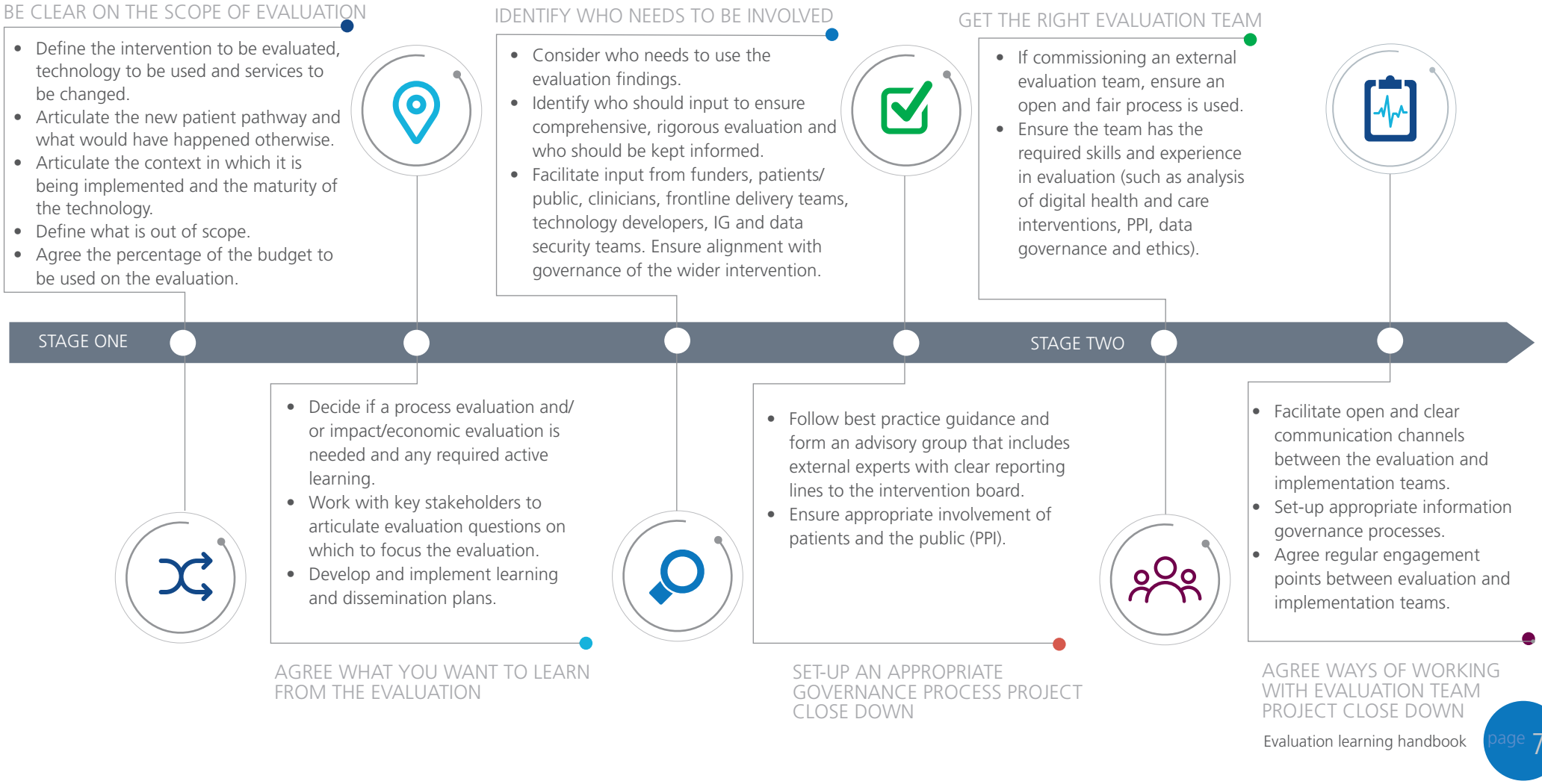
Diaries

Participants keep diaries about their activities and experiences. This can be structured (recording quantity and using time slides) or provide a narrative (writing up events).

Wave 1 Test Beds' approach to evaluation

Each Test Bed local evaluation team generated evidence for their site individually; however, it is also important to learn about common themes from across Test Beds which have potential for wider application. The Test Beds National Evaluation Partner played an important role in synthesising local Test Bed findings and identifying shared challenges and enablers that can help inform national practice and the work of future Test Beds. Based on Wave 1 experience, it is vital to plan for and address evaluation risks through the life of a project – from preparing to commission an evaluation, selecting the evaluation team, being an intelligent customer for the evaluation, and learning and disseminating findings. In line with national best practice, figure 1 below outlines some of the key activities conducted by Test Beds and key learning reflections from Wave 1 for each of these stages, including when is most helpful to carry them out. Please note that the timings are indicative and subject to local circumstances (e.g. resources and timescales).

Figure1: the Test Beds evaluation journey



STAGE THREE

- Ensure a logic model framework for the evaluation is prepared.
- Ensure the most robust methods are selected, given data and practical constraints.
- Ensure evaluation methods are appropriate, given the maturity and context of the intervention.
- Ensure selected evaluation methods can be integrated to maximise learning.

ENSURE JOINT OWNERSHIP OF EVALUATION APPROACHES

BE AN INTELLIGENT CUSTOMER FOR THE EVALUATION

- Play a scrutiny role. Ask questions of the evaluation team to ensure you fully understand the activities underway.
- Monitor patient recruitment activity and be prepared to suggest adapting to maximise numbers.
- Test and challenge findings to ensure appropriate interpretation, including about distributional effects.

ENSURE PREPARATION OF A CLEAR EVALUATION PROTOCOL

- Encourage evaluators to co-design an evaluation protocol with the implementation team.
- Agree accountabilities across evaluators and implementers.
- Ensure evaluators obtain relevant ethics clearance on evaluation protocol (e.g from the HRA).
- Keep protocol updated over time.

ENSURE FINDINGS AND CONTEXT ARE CLEAR IN FINAL REPORTING

- Ensure context for the interventions is clearly conveyed i.e. the system into which the intervention was implemented.
- Ensure findings identify the conditions under which outcomes were observed, for whom and why.
- Sense-check results and findings, including limitations and unintended consequences.

STAGE FOUR

- Identify key points of learning from the evaluation.
- Identify practical ways to put learning into action for future interventions.
- Translate evaluation findings for policy makers to use.
- Set up framework to continue to monitor and evaluate on-going interventions.
- Make appropriate learning experiences available to relevant stakeholders.

TAKE STOCK OF LEARNING AND EMBED IT IN FUTURE ACTION

ENSURE DISSEMINATION PLAN IS PUT INTO PRACTICE

- Continue appropriate sharing of learning throughout the evaluation.
- Ensure dissemination materials are prepared with clear messages, in line with the learning and dissemination plan.
- Ensure frontline staff, patients and public are among those with whom learning is shared.

Reflections from the National Test Beds Team

Having an independent evaluation in place is a key condition of funding for Test Beds. The National Test Beds Team values evaluation as it generates evidence about what works. It allows us to learn about how well health and care interventions have been implemented and what could be done better; the extent to which they have delivered the targeted outcomes and for whom, and why (or why not) and under what conditions; and importantly, whether they have delivered value for money.

Independent evaluation plays an important role in raising issues that other people might not raise. By retaining its independence, evaluation will have the greatest value.

As part of the set-up of Wave 1 of the programme, Test Bed local evaluators were required to conduct two types of evaluation activity which focused on different aspects of the interventions that sites were testing:

- Process evaluation: this explores how the intervention(s) work and why, if the implementation of the intervention has been successful, what changes were made during implementation and why, and identifies the barriers and facilitators to delivery.
- Impact and economic evaluation: this explores the extent to which clinical outcomes and patient experience have been improved at the same or lower cost of service delivery to the NHS.

A full list of the evaluation questions asked and answered by evaluators in Wave 1, and which will inform evaluation in future waves of the programme, are explored in section 2 overleaf.

Key learning reflections

Three key clear high level themes arose from the evaluation processes followed by sites in Wave 1:

- There is no one-size fits all method for undertaking an evaluation. In Wave 1, each of the Test Beds used different evaluation methods to assess the impact and outcomes of their interventions because it is important that the evaluation method is appropriate for the particular characteristics of the intervention. The maturity or 'stability' of the intervention is very important i.e. whether the technology is continuing to be adapted to match the patient needs, or if it is stable over time. A range of different quantitative methods were used. For example, the Technology Integrated Health Management Test Bed used a randomised control trial to assess the impact of technology that supports patients with dementia, whilst the Heywood, Middleton and Rochdale Test Bed compared secondary care activity in the area in which the intervention was implemented with secondary care activity in a different area that experienced the same pattern of change in that activity over recent years (a difference in differences of secondary care data). It was clear that where the Test Beds, and local evaluators, selected mixed methods (combining qualitative and quantitative approaches), they were better able to respond to the innovations and range of interventions being tested and also the evaluation aims and objectives identified by local providers, commissioners, service users and carers.
- It is essential to allow time and involve all relevant stakeholders in the evaluation right at the start to agree on the evaluation objectives; focus of the analysis; and how data will be collected. Accepting that not every wish of every stakeholder can realistically be included.
- Getting the right people around the table at the start is essential. Not only does this mean every relevant stakeholder is sighted on what the evaluation will look like and how it will work, but it ensures that maximum learning can be generated from the evaluation therefore maximising the efficiency of evaluation resources. Key stakeholder groups to ensure representation of at the start are: patients and service users; clinicians; innovator partners; health and care service delivery leaders; local and national evaluators; the Test Bed leadership team; and Test Bed funders among others.
- There is a need to consider qualitative and quantitative evidence together: this is essential to formulate the comprehensive narrative about the intervention. Qualitative research gives insights into why and how things happen, while quantitative research provides more confident estimates of the extent of change, causes and effects and differences across contexts and subgroups. Hence all good evaluation will combine both approaches.

Useful links

Toolkit for communications and engagement teams in service change programmes: https://improvement.nhs.uk/resources/communications-and-engagement-toolkit-teams-service-change-programmes/

Standards of Evidence – Clinks Guide: https://www.clinks.org/sites/default/files/StandardsofEvidenceGuide.pdf

Better Evaluation: http://betterevaluation.org/sites/default/files/Rainbow%20Framework%20-%20compact%20version.pdf


Section 02

Preparing to commission an evaluation

Learning from the set-up of Wave 1 Test Beds has highlighted four key critical steps that can support future Test Bed Programme Directors and project staff when preparing to commission an evaluation:

Clarifying where possible the intervention to be evaluated

Being clear about who needs to be involved

Being realistic when establishing what you and others want to learn from the evaluation; establish the questions to be answered.

Exploring practical considerations

- Resources (time and money)
- Evaluation governance
- Ethical approval
- Data collection and sharing

Useful link

Further information to consider when preparing to commission an evaluation can be found here: https://mrc.ukri.org/documents/pdf/mrc-phsrn-process-evaluation-guidance-final/

Clarifying the intervention to be evaluated

Context

Wave 1 Test Beds learnt that to be able to evaluate an intervention, clarity is essential about the nature and scope of the intervention being implemented, and in particular, how the patient pathway of care is likely to change, relative to what would happen without the intervention. Therefore both the technology being introduced and the changes to the wider health and care system must be mapped out. The rationale behind it must be articulated, along with how it is supposed to work as this will affect the appropriate evaluation design.

Reflections from Wave 1 of the Test Beds programme

As part of the set-up phase of Wave 1, Test Bed sites were expected to develop a Theory of Change or logic model which described their intervention:

- A Theory of Change is a comprehensive description of how and why a desired change (or outcome/impact) comes about from an intervention.
- A logic model is a framework which specifies the inputs (people, resources, IT etc.) required to deliver and operate an intervention; the activities these inputs facilitate; the product or service outputs which are tangibly delivered by the intervention; the outcomes for the patients or users, carers, staff (or others), which in turn lead to longer term impacts.

Not having a clearly defined logic model will make it hard to identify the metrics against which progress and performance of the intervention can be assessed.

Learning recommendations

Based on Wave 1 Test Beds experience, a future wave of Test Beds should consider and clearly set out the following early on and before delivery begins:

1. The objectives of the intervention (e.g. improved outcomes for a defined group of patients or service users; efficiency savings etc).

2. A description of the intervention, including:

- Where the intervention will take place (e.g. geographical locations or services, including whether they cross organisational boundaries or primary and secondary care).
- The activities involved in the intervention.
- How activities are expected to link together to lead to the intended outcomes.
- Potential unintended outcomes that should be monitored.

3. How novel or innovative the intervention is.

- Is there an existing evidence base underpinning the intervention, or is it completely new and untested?
- Is the intervention expected to change and evolve significantly as it is implemented, or is it expected to be relatively stable? Experience from Wave 1 Test Beds and similar programmes suggest that interventions are likely to be highly innovative and will change over time.

This will need to be shared with prospective evaluators as part of the evaluation tendering/commissioning process (see Section 3) and the stakeholders.

Useful links

Logic modelling: WK Kellogg Foundation, Logic Model Development Guide

Magenta Book, HM Treasury, 2011

Be clear about who needs to be involved

Context

Wave 1 Test Beds learnt that getting the right people involved at the right time has a substantial impact on how straight forward it is to take forward an evaluation; extracting what needs to be learned to add maximum value; and how robust the analysis will be.

Reflections from Wave 1 of the Test Beds programme

Wave 1 of the Test Beds programme centred on collaboration. Test Beds sites involved clinicians and other health and social care professionals, operational staff and managers from different organisations, including Academic Health Science Networks (AHSNs), innovators and third sector organisations, working together in new ways for the first time. In addition, the partnerships involved non-traditional partners such as small and medium sized firms (SMEs) and industry representatives. Co-production with patients, users and carers was also an important part of the programme and users were involved to varying degrees in the selection of interventions, governance and evaluation.

Wave 1 Test Beds had formalised governance and management structures, typically in the form of a Programme Board. 'Sitting beneath' the overarching programme board were a series of work stream leaders and each Test Bed had an advisory group or steering group that focused on evaluation. Sites' evaluation groups typically consisted of the following stakeholders: patient/ service user representatives; local evaluators; National Evaluation Partners; innovators; and programme managers.

A number of sites had public and patient involvement (PPI) as part of their governance groups. For example, the Diabetes Digital Coach Test Bed in the West of England had a patient representative who has diabetes sit on their Evaluation Advisory Group. The Perfect Patient Pathway (Sheffield) Test Bed had a local Healthwatch representative sitting on their evaluation advisory group who engaged with the wider public, ensuring that any views were incorporated into the delivery of the programme including evaluation.

PPI is an increasingly important dimension of research and evaluation as it can improve the quality of evaluation and resulting findings as well as support ethical considerations. The nature of PPI ranges from light-touch involvement through to co-production and patient-as-researcher models. Who is considered the public or patient also varies – from working with a small number of 'expert' patients or patient representative organisations, to involvement of a larger number of regular service users or the general public. Different approaches are appropriate for different contexts. What all approaches have in common is that PPI is about more than just including patients as respondents in the research (e.g. by interviewing them).

Learning recommendations

Based on Wave 1 Test Beds experience, it would be helpful for a future wave of Test Beds to:

Consider, as early on as possible, who needs to be involved in determining how the evaluation will be conducted; what kind of information it will generate and how that information will be shared, with whom and when.

Key people to include might be:

- The 'sponsors' of the intervention being implemented (e.g. intervention funders; leaders/managers of organisations in which the intervention is being implemented; leaders of innovation partners).
- Intervention designers and implementers (e.g. intervention project director; intervention project manager; members of the implementation team; technology partner organisations; or, voluntary sector organisations) and people involved in its delivery (e.g. clinicians, managers, administrators and allied health professionals involved in wider implementation).
- Representatives of service users or patients that will be using or affected by the intervention (e.g. Healthwatch or a relevant patient charity).
- Data and IG experts from the Clinical Commissioning Group (CCG)/ Trust/ General Practitioners (GPs) etc. who will need to be relied upon for essential data.

Based on the Wave 1 experience, it is also helpful for Test Bed Programme Directors and project staff to:

- Map out who they want to involve in the evaluation and why. This could be the responsibility of programme leads or could be led by the evaluation team. A good evaluation team will ensure that all relevant stakeholders are engaged appropriately.
- Scope what level of involvement different stakeholders will need in the evaluation and when. Some groups will need to be closely involved whilst others will only have a broad interest in its findings. Successful Wave 1 Test Beds found that it was helpful to think about the best way of getting each stakeholder group involved at all relevant stages of the evaluation process, before, during, and after its commission.
- Get stakeholder groups on-board as early as possible, and allow them to make as much contribution to evaluation design as able within the constraints of what is appropriate.
- Be mindful that the evaluation may find that the intervention is not successful in achieving its aims or not good value for money. Although this may be an important and valuable finding, it can be disappointing to stakeholders, therefore managing expectations throughout the evaluation process is key.

Establish what you and others want to learn from the evaluation

Context

Wave 1 Test Beds learnt that it is important to be clear on what they wanted to learn about to help make sure that the evaluation is well-targeted and can focus on the issues that matter most.

Reflections from Wave 1 of the Test Beds programme

Understanding what the relevant stakeholders want to learn from the evaluation at the start is essential. Some mapped this out clearly at the start, whereas others found that the work was being pulled in different directions as consensus was not reached early enough about what to explore in the evaluation and how best to do it. In the latter case, this led to time and effort being required to narrow and refocus the scope of the evaluation.

Striking a balance across the views of different stakeholders can be challenging, especially with a fixed set of evaluation resources. Strong leadership by the evaluation team leader and the programme director are therefore needed to explore and unearth views from relevant stakeholders as early as possible, and to articulate what will and will not be in scope of the evaluation.

Learning recommendations

Based on Wave 1 Test Beds experience, it would be helpful for a future wave of Test Beds to:

- Think about the overarching purpose of the evaluation.
- Identify the different aims and priorities for key Test Bed stakeholders and what they want to learn from the evaluation.
- Establish a consensus regarding the aims and priorities for the project based on the needs of key stakeholders, so that the evaluation is focused and can be undertaken within available resources to a good standard.

Think about the aims and needs of future funders to support the commissioning, spread and scale of successful interventions:

- Be sure to explore with them at an early stage what form of evidence they would want from an evaluation to support their funding decisions.
- Include questions to address what they would consider to be a robust evaluation approach.
- Be aware that this process may take time, and this should be built into the timeline for commissioning the evaluation.

Think through process evaluation questions and impact / economic evaluation questions, for example:

Process evaluation questions

- What is the Theory of Change for the intervention: how is it supposed to work and why?
- How was the intervention, and the partnership involved in the intervention, set-up, designed and established? What can be learnt about improving this process?
- Did the partnership of the NHS and other organisations work and why? Has this partnership/different engagement with the NHS resulted in improved technology pull-through? What could be improved?
- Was the intervention delivered in line with the original plans? Were changes made to the design of the intervention, and if so why? What changes had to be made to implementation plans to ensure effective delivery of the intervention, and why?
- What were the barriers to effective delivery (and uptake of technology/ services) and how were they overcome?
- What were the facilitators of effective delivery (and uptake of technology/ services) and how were they ensured?
- Were there any unintended consequences that needed to be managed and how was this done?
- To what extent is the intervention likely to be scalable and why? If the intervention isn't scalable, what are the specific barriers to prevent it being adopted and spread to other sites.

Impact evaluation questions

- What was the impact of the intervention on patient experience and how did this differ across groups of patients, and why?
- What was the effect of the intervention on clinical outcomes and did this vary between groups of patients?
- What were the cost implications of the intervention from the NHS perspective and did it constitute value for money?
- What are the conditions under which the intervention was most effective, for whom and why

Consider if there are any additional questions and/or can they be more specifically tailored to the intervention

Context

Wave 1 Test Beds found that the following practical considerations had a substantive impact on the evaluation of the interventions.

Resources (time and money)

Reflections from Wave 1 of the Test Beds programme

Time and money were critical success factors in supporting the evaluation of Wave 1 Test Beds and impacted on different aspects of Test Beds' evaluations in several ways:

Time

- Recruitment: several Test Beds were ambitious in the numbers of patients that they could recruit to test their intervention in the time available. For example, the Technology Integrated Health Management Test Bed wanted to recruit 700 pairs of patients and their carers (dyads); in the end 102 dyads were recruited.
- Test Bed implementation: most Test Beds took notably longer to implement their interventions than intended. This had a detrimental impact on the ability to undertake evaluation e.g. RAIDPlus were not able to undertake any impact evaluation because by the end of the programme they still were not in a position to go live. Several other Test Beds had to shorten the duration of running the Test Bed interventions.

Money

Budget allocation to evaluation resources must be proportionate. It is important to be mindful that resources must be sufficient to fund an evaluation team over the full duration of the programme (2 years). In some cases, insufficient resources were allocated to evaluation which needed remedial action.

Learning recommendations

Based on Wave 1 Test Beds experience, it is important that a future wave of Test Beds allocates time to:

Consider the timing of the evaluation

- Agree the period over which the intervention will be implemented and monitored; the related timeframe of the evaluation; and what evaluation questions can be realistically answered within these timeframes.

Think about

- The short-term and long-term outcomes of the intervention: If the expected ultimate impact of the intervention may not be available for a long time, think about what intermediate outcomes would be expected to be seen in the shorter-term and what near-term outcomes could provide confidence that the intervention is likely to have the anticipated impacts over time.
- The important decision points of the team, funders, and innovation partners– identify and map these and establish what information from the evaluation needs to be available when.

Useful link

"In determining the amount required to finance the evaluation function, other organizations have estimated that 3–5% of the programme budget should be used for evaluation." WHO (2013), Evaluation Practice Handbook.

Decide on the budget of the evaluation early on

- Ensure that adequate budget is available to undertake a robust evaluation. Test Beds evaluation budgets need to be appropriate for the scale and duration of the evaluation, and the number of evaluation questions explored.

Data sharing and data collection

Context

Information and data sharing is at the heart of many Test Beds innovations, but this must always be in full compliance with relevant regulations such as the General Data Protection Regulation (GDPR).

Reflections from Wave 1 of the Test Beds programme

In Wave 1, information governance was a critical issue. Some local evaluation teams required access to sensitive data such as patient reported outcomes; patient-level administrative data; and patient-level data automatically collected online via the technology being implemented. Several Test Beds struggled to access data in a timely way, including from consortium partners due to the need to sign up to data sharing agreements. As part of Wave 1 of the programme, Test Beds were required to think about the information governance implications of their projects. We encourage future Test Beds to refer to the Test Beds information governance learning handbook for further information:

Learning recommendations

Based on Wave 1 Test Beds experience, it may be helpful for a future wave of Test Beds to early on in the project:

- Bring together data and IG experts from the Clinical Commissioning Group (CCG)/ Trust/ General Practitioners (GPs) etc. who will need to be relied upon for essential data.
- Agree and articulate a clear process for data ownership, processing and sharing in line with relevant regulations (such as GDPR).
- Explore the most appropriate data sources, engage data owners to establish what approvals are required for access to the data to be granted, and to apply and seek approval for data to be released.
- Consider evaluation and access to data as part of the setting up of Test Bed partnerships and the design and delivery of the intervention.
- Agree who holds the responsibility for collecting the data and the method for collecting it.

Useful link

NHS England (2018) Test Beds: Information Governance Learning from Wave 1: https://www.england.nhs.uk/publication/nhs-england-test-beds-programme-information-governance-learning-from-wave-1/

Ethical approval

Context

It is important to be aware that ethical approval may be required for the evaluation. A future wave of Test Beds is likely to require ethical governance and permissions for the intervention, and this is also true of the evaluation.

Reflections from Wave 1 of the Test Beds programme

Ethical approval was appropriately considered by all Wave 1 Test Beds. Most did not require formal approval from the Health Research Authority (HRA) because their evaluations were service evaluations and not research. For example, following discussion with the HRA and the Research and Development Directors of the NHS Trusts involved in the Lancashire and Cumbria Innovation Alliance programme, it was agreed that phase 1 of this intervention was service evaluation and as such did not require HRA approvals. Phase 2 however was deemed by the HRA to be gathering new data and was thus defined as 'research', therefore requiring HRA approval:

- **Most Wave 1 Test Beds** undertook a service evaluation so did not need HRA clearance. They did however, require the clearance of their university ethics committee.
- **RAIDPlus** sought local and national ethics approval for their project which was not factored into their initial plans and found that in practice this was time consuming and impacted on their overall delivery timelines.

Learning recommendations

Based on Wave 1 Test Beds experience, it is important that a future wave of Test Beds:

- Factor in obtaining ethical approval as part of the overall project planning process for their project – gaining ethical approval can be a relatively lengthy process, so this needs to be factored into evaluation timescales.
- Use the HRA tool as a guide through the various forms of approval which may be required (http://www.hra-decisiontools.org.uk/research/). The nature of the evaluation and what it involves will affect what kind of approval is needed. For example, 'research' requires HRA approval whereas 'service evaluation' does not.
- Let the National Test Beds Team know if they need support in applying for HRA approval early on. The National Test Beds Team has strong links with the HRA and can provide support.

Useful link

https://bmjopen.bmj.com/content/bmjopen/8/2/e017268.full.pdf

Evaluation governance arrangements

Context

Wave 1 Test Beds learnt that good governance is essential to ensure that all involved can be confident that the evaluation is independent, objective and robust.

Reflections from Wave 1 of the Test Beds programme

Each Wave 1 Test Bed had a formalised governance and management structure comprising an overarching or umbrella programme team or board which oversaw the work of an evaluation advisory group and/or local evaluation teams. As part of the set-up process, Test Beds reflected on the governance arrangements for the evaluation, and how these relate to, and interact with, governance arrangements for the intervention. They considered:

- The formal reporting requirements of different stakeholders (particularly funders and the Board(s) of the organisation(s) in which the intervention is being implemented).
- The level and nature of patient and public involvement (PPI) that was needed and appropriate for the evaluation.
- Whether establishing an advisory group for the evaluation was required (best practice would be to draw upon independent experts to provide advice and quality assurance).
- Whether ethical clearance was required.

Learning recommendations

Based on Wave 1 Test Beds' experience, it may be helpful for a future wave of Test Beds to also consider the following issues before establishing an advisory group:

- The terms of reference including whether the group is purely there to offer advice or whether it has some degree of influence over the evaluation.
- What the group needs to achieve. For example, to provide expert methodological advice; to provide expert substantive knowledge to the evaluation relevant to the focus of the intervention; to offer practical advice around common challenges such as recruitment.
- Who is on the group and why. Consider which stakeholder groups should be included and ensure that at least some members have evaluation expertise. Advisory groups ideally include people who are unconnected to the intervention or evaluation but who have relevant expertise, including in evaluation methodology.

Useful link

Evaluation governance arrangements Magenta Guide Table 5C, annex 2.0


Section 03

Commissioning an evaluation

Wave 1 Test Beds commissioned a range of organisations to conduct their local evaluations.

Ensuring a fair and open procurement process

Reflections from Wave 1 of the Test Bed Programme

Test Bed teams initially reported that procurement of an external evaluation team felt purely like an administrative process, but on reflection realised that commissioning an independent external evaluation team was actually an important part of the evaluation timeline.

During the procurement process Test Bed teams invited suitably skilled external parties to submit their ideas on how to carry out the evaluation, and to demonstrate why they would be the best evaluation team to meet the needs of the Test Bed in terms of quality and budget. Test Beds reported that they had to consider that they were selecting one of these evaluation teams to work closely with them.

Learning recommendations

Based on Wave 1 Test Beds experience, it may be helpful for a future wave of Test Beds to:

- Engage the procurement team early on in the process to establish how the evaluation team will be selected and contracted. For example, via single source, mini-, open-competition. This process should be fair, transparent and consistent with procurement rules.
- Prepare a specification, which should provide sufficient information for potential bidders to have a good enough understanding about the intervention; what is required from the evaluation team in terms of delivery, skills and experience; the timelines they must adhere to; and the working arrangements they are to follow.

Appointing an evaluation team

Reflections from Wave 1 of the Test Bed Programme

Test Bed sites from Wave 1 used interviews to select an evaluation team from amongst the tenders received and stressed the value of investing the time to ensure that they made an informed decision about the appointment.

Learning recommendations

Based on Wave 1 Test Beds experience, it may be helpful for a future wave of Test Beds to:

- Convene a panel to carry out the interviews with approximately 3-5 people, including a procurement representative. The Wave 1 Test Bed sites agreed that members of the panel should be clear about what they want to achieve in the evaluation, and have sufficient knowledge to be able to objectively probe particular aspects of the tender response. The panel may include clinical expertise, evaluation expertise, and public and patient involvement.
- Interview questions should always cover the main elements of the proposal (e.g. methodology, team, proposed outputs, risks, value for money), and allow the panel to explore the key questions of interest. The approach to interviews should be systematic, to ensure fairness, and agreed with the procurement team in advance.
- Continue to engage with the procurement team on appointment of an evaluation team so they will be able to advise on contractual arrangements, payment milestones and other legal matters.

Agreeing how the team will work together

Reflections from Wave 1 of the Test Bed Programme

Wave 1 sites agreed that establishing a process by which they worked with the evaluation team was essential. They felt it was important to work collaboratively through a number of key areas including: information governance, intellectual property (IP) and day-to-day working.

Learning recommendations

Based on Wave 1 Test Beds experience, it may be helpful for a future wave of Test Beds to:

- Agree contractually who owns the IP as the program will be generating new information and knowledge about the impacts of the evaluation. This is often handled through standard contracting procedures.
- Articulate a clear process around information governance (IG) to ensure that data ownership, processing and sharing meets relevant regulations such as GDPR. Wave 1 sites felt that even though IG issues may have arisen during the initial partnership set up, they should also be considered in the context of evaluation. For example, what data will be required, where it will be sourced, what IG procedures will be followed, any likely IG challenges (e.g. analysing patient-level data).
- Establish how the team work together to ensure that work flows efficiently and effectively on a day to day basis. Wave 1 sites used a Project Initiation Document (PID) to describe how often they would communicate with the evaluation team and what should be shared and when as well as escalation procedures for when work wasn't going to plan. PIDs are also helpful for defining: [1] what is in scope vs out of scope, [2] individual responsibilities, [3] what will be required from you, the client and [4] expected risks and mitigation measures.


Section 04

Being an intelligent customer for a robust evaluation

Agreeing the evaluation approach

Reflections from Wave 1 of the Test Bed Programme

Wave 1 sites recognised a core responsibility to ensure that the evaluation delivered the type of evidence that they needed and that it was of a suitably high standard, requiring them to be an 'intelligent customer'. This meant that they asked the right probing questions at the right time and had the right checks and balances in place so that they were confident about the quality of evidence they received. This may have included an expert advisory panel overseeing the work.

Learning Recommendations

Based on Wave 1 Test Beds experience, it may be helpful for a future wave of Test Beds to:

- Discuss and agree the evaluation approach early on with the evaluation team. The proposed evaluation approach should be discussed, refined building upon the proposal/ITT stage and agreed so that there are clear and shared expectations about the scope of the work; what is needed to make it happen (such as data required, input from others etc.) and what the deliverables will be throughout the evaluation (such as an interim report; Board papers; and final report).
- Think about the important components of an evaluation: the logic model framework; categories of evaluation; types of evaluation method; and agreeing the evaluation approach.

Developing logic models

Reflections from Wave 1 of the Test Bed Programme

The Wave 1 sites found that evaluation activity began with a clear framework that allowed them to set out the objectives of the intervention, along with the mechanisms through which the intervention was expected to deliver the desired outcomes. By mapping this out clearly and concisely it allowed them to think hard about what factors need to be in place for outcomes to be delivered. A logic model was the framework often used to map out the intervention in this way.

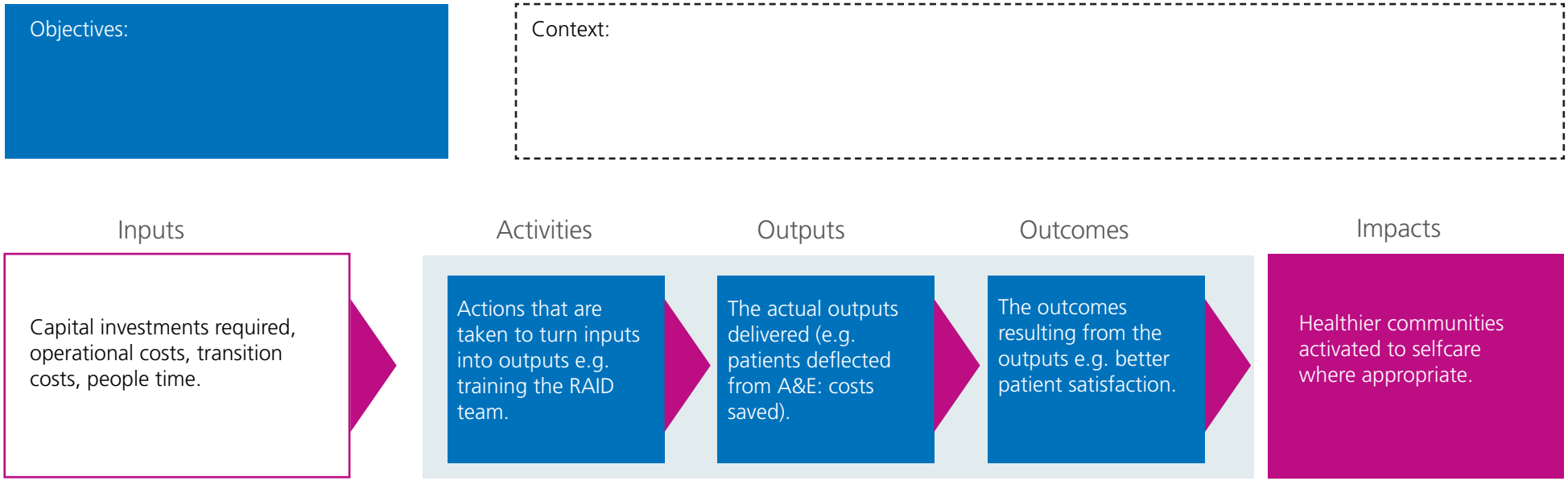
More specifically, a logic model specifies the inputs (people, resources, IT etc.) required to deliver and operate the intervention. These inputs are used to deliver a number of activities which in turn are intended to deliver a number of product or service outputs. The activities and outputs should result in beneficial outcomes for the patients or users (or others), which in turn lead to longer term impacts. A suggested logic model structure is illustrated in Figure 2 and an example of a logic model used by a Test Bed is included in annex 1.2

Learning Recommendations

Based on Wave 1 Test Beds experience, it may be helpful for a future wave of Test Beds to:

- Ensure that a logic model is done, taken seriously, and is agreed by a range of those involved to ensure it is accurate.

Figure 2: the logic model



The logic model helps both the process evaluation by clarifying what is to be implemented (i.e. the inputs, activities and outputs) and the impact/ economic evaluation by helping them to identify the indicators or metrics they could use to measure the extent to which the intervention has delivered the desired outcomes, or has had other unintended effects. This in turn helped to identify which data needed to be collected.[1]

Useful link

More information on logic models can be found at: logic modelling: WK Kellogg Foundation, Logic Model Development Guide: https://www.wkkf.org/resource-directory/resource/2006/02/wk-kellogg-foundation-logic-model-development-guide

1 More information and examples of logic models for complex interventions can be found at https://www.strategyunitwm.nhs.uk/publications/logic-models-complex-programmes

Understanding the selection of evaluation methods

Reflections from Wave 1 of the Test Bed Programme

As noted in Section 1, Wave 1 sites used a variety of evaluation methods, which included both process and impact evaluations, to assess the impact of their interventions. The methods that they used were determined by a range of local factors such as data availability, budgets available, timeframe for the analysis and active consideration of the maturity of the intervention.

Although the evaluation team select the methodology that is used in each Test Bed site, it is important for Programme Directors to be involved in discussions around the reasons for selecting particular methods and ensure they have a good understanding of evaluation issues.

Test Beds found that the quality of evidence generated varied with each method and that different evaluation approaches presented advantages and challenges. For example, if one simply compares outcomes before the intervention with outcomes after, this ignores the possibility that other factors could have significantly influenced the change observed (i.e. there is a risk of wrongly attributing any changes to the intervention). Randomised control trials on the other hand have traditionally been considered the most robust – but they are not appropriate in all circumstances, e.g. in developmental projects when the intervention is being refined; also randomised control trial evaluations can take longer to plan and undertake than other forms of evaluation and therefore may not be feasible in practical terms. There is therefore a decision to be made about which method is fit for purpose and strikes an appropriate balance between simplicity and resources, and rigour.

Learning Recommendations:

Based on Wave 1 Test Beds experience, it may be helpful for a future wave of Test Beds to:

- Consider the individual characteristics of both their intervention and the evaluation, and the contexts under which they are being undertaken before selecting evaluation methods.
- Press their evaluation team to consider various options and be clear on the justification for any preferred option.
- Work with the local evaluation team to produce a well-structured evaluation protocol which covers all aspects of the evaluation, both qualitative and quantitative. The purpose of a protocol is to have a record that can be shared with stakeholders and sets out the focus of the evaluation; what analysis will be carried out; what (e.g. data) and who (e.g. qualitative sample frame) is required to be involved; risks and how they will be mitigated; timelines; and designation of responsibilities.

Learn more about evaluation methodologies:

- Process evaluation – logic model/theory of change workshops, logic model/theory of change building exercises, key stakeholder interviews, stakeholder interviews, observation and for mature interventions only – assessment of fidelity.
- Impact evaluation – before vs after, before vs after with controls, econometrics, quasi-experimental methods and experimental methods – randomised control trials. Refer to the supporting information found in Annex 2.0 – 5.0 and the useful links below. Common challenges and key issues for consideration.

Useful links

http://www.whatworksgrowth.org/resources/the-scientific-maryland-scale/

https://mrc.ukri.org/documents/pdf/mrc-phsrn-process-evaluation-guidance-final/

Technology Integrated Health Management – Randomised control trial

The Technology Integrated Health Management Test Bed led by Surrey & Borders NHS Foundation Trust, conducted an exploratory randomised controlled trial in one area (Surrey and North East Hampshire region) to explore the outcomes of six months' use of technology integrated health management for Dementia – a health monitoring and managing system – for those with early/moderate dementia. The trial included a health economic element (to understand Technology Integrated Health Management for Dementia's costs and benefits). It also included a process evaluation (to understand factors in the workings of the Test Bed – and intervention itself – that helped or hindered progress).

The evaluation intended initially to conduct a definitive randomised controlled trial. However, due to constraints of time on recruitment and the maturity of the intervention, it was determined that an exploratory trial which allowed exploration of outcomes rather than definitive conclusions to be reached was more befitting. This decision was reached in conjunction with the funders.

Care City – Simulated modelling

The Care City Test Bed used modelling to assess the impact of Kardia Mobile™ in the absence of actual cost and impact data. Kardia Mobile™ is a device for use with smart phones that helps to detect atrial fibrillation (an irregular heartbeat). Their local evaluation team constructed a model to estimate the impact on costs and outcomes of a programme of annual screening for atrial fibrillation with Kardia Mobile™ in a community pharmacy setting with onward referral to a dedicated clinic. The model was populated using evidence from the literature relating to atrial fibrillation, the attendant stroke risk, the cost and impact of strokes and the cost and effectiveness of anticoagulation to reduce stroke risk. Data from the Care City pilot was used to estimate the cost of the pathway as implemented.

Common challenges and key issues for consideration

Timely access to required data

Reflections from Wave 1 of the Test Beds programme

Data was an essential ingredient in the evaluations conducted by Wave 1 Test Beds. However, at various points in the programme several sites encountered difficulties in accessing appropriate data sources and it took a large amount of time. For example, the Heywood, Middleton and Rochdale (HMR) Test Bed encountered difficulties with access to secondary care data nationally and regionally. Care City's local evaluation team similarly experienced delays in obtaining data from a local clinical commissioning group.

Learning recommendations

Based on Wave 1 Test Beds experience, it is important that a future wave of Test Beds considers having access to high quality data sources early on in their projects and choose data sources wisely. Figure 3 lists examples of existing data that can support evaluation.

Figure 3: Illustrative types of existing data which can support evaluation

Type	Examples	Useful for	Sources
Existing administrative data	Hospital Episode Statistics, Reference Costs	Comparators / benchmarks, estimating activity and financial impacts	NHS England, Department of Health, NHS Digital, Office for National Statistics
Secondary/ admin survey data	Inpatient Survey, GP Patient Survey, NHS Staff Survey, Friends and Family Test	Comparators / benchmarks, estimating quality impacts	NHS England, NHS Digital, Office for National Statistics
Intervention monitoring data	Intervention costs, activity rates, patient characteristics, clinical outcomes	Estimating clinical and economic impact of intervention	Local services / intervention deliverers

Patient recruitment and retention

Reflections from Wave 1 of the Test Beds programme

Test Beds experienced challenges with patient recruitment – this is an issue for evaluation as it drives sample sizes from which we can draw conclusions, and fewer observations = less statistical confidence in the findings. Several Test Beds had small sample sizes and some experienced attrition – patients dropping out of their evaluation. Recruitment to local interventions was challenging as it often involved obtaining formal consent from each individual patient and in some cases patients were nervous about engaging with a new service. Test Beds responded in a creative manner and recruited patients in a range of ways including:

- Recruitment via third parties such as GPs and pharmacies who have direct contact with patients.
- Conducting recruitment campaigns involving advertisements in the local press and stands at events.
- Inviting patients to self-refer into the programme.
- Working in partnership with patient and carer groups and voluntary and community sector organisations.

Learning recommendations

Based on Wave 1 Test Beds experience, it may be helpful for a future wave of Test Beds to employ the following strategies to help attract, recruit and retain patients.

Focus on recruiters:

- Develop relationships with local practitioners, so that they understand the value of local interventions and how they can benefit patients, and create mechanisms for physician referrals into the scheme.
- Work with patient/care representatives and voluntary and community sector groups to help promote and publicise work
- Set realistic targets for patient recruitment within the time they have available, ensuring this aligns with the evaluation and outcomes they hope to achieve.

Maintain engagement with patients and carers that are involved:

- Produce newsletters and/or attend meetings with local voluntary and community sector groups to help communicate the benefits of being involved in the evaluation and where appropriate, communicate early findings.
- Recruit patients continuously.

Communication flows between evaluators, implementers and other key partners

Reflections from Wave 1 of the Test Beds programme

The interventions being tested by Test Beds locally included a range of innovations and multiple stakeholders including practitioners, patients and technology partners. At times, Test Beds found that communication channels broke down between particular groups and they did not work as effectively. Confusion about responsibility and accountability at times led to conflict between partners in terms of how activities should be carried out and delays which risk insufficient time for data collection and analysis of evaluation evidence.

Learning recommendations

Based on Wave 1 Test Beds experience, it may be helpful for a future wave of Test Beds to:

- Establish the governance structure for the project as a whole (the intervention and the evaluation) at the outset and monitor that processes are implemented to ensure all partners are working together effectively and that communication channels are clear and working well.
- Keep open channels with technology partners between all stakeholders, but especially between evaluators and programme directors, evaluators and tech partners and evaluators and data providers.

Interpretation of results

Reflections from Wave 1 of the Test Beds programme

Wave 1 Programme Directors were responsible for commissioning local evaluations and ensuring the evaluation evidence was robust, clearly presented and able to stand up to scrutiny. As part of their quality assurance work, they asked local evaluators appropriately challenging questions and made sure that they themselves understood the findings and conclusions that had been drawn. Programme Directors also tried to ensure there was sufficient transparency in written documents – making sure that external parties could understand the underlying evidence and assumptions that underpinned the work. For example, understanding the counterfactual (what would have been expected if the intervention had not been implemented) and the research/data used to inform economic modelling in local sites.

Learning recommendations

Based on Wave 1 Test Beds experience, it may be helpful for a future wave of Test Beds to similarly play a scrutiny role in overseeing local evaluations and probe in to the following particular areas (see Annex 6.0 for detailed explanation):

- Strength of findings on process and drawing on qualitative methods[2]
- Strength of findings on impacts
- Headline findings
- Unintended consequences
- Scaling up

Reflections from Wave 1 of the Test Beds programme

2 Just like quantitative methods, there are quality standards which qualitative methods should adhere to in order that findings generated are reliable.

Presentation of emerging results and preparation of final reports

Based on the experiences of Wave 1 Test Beds, it may be useful for future waves to use a dashboard to compile and present key information and evidence in an impactful and easy to understand way. Figure 4 is an example dashboard which includes suggestions of information that Programme Directors and decision-makers may find helpful. Where common metrics are included it would be useful to provide qualitative information to explain them and help make sure that they can be appropriately interpreted. The metrics and dashboard may not be feasible or appropriate for some interventions; however, it could be tailored where helpful.

Each local evaluation team was required to produce a checkpoint report, sharing emerging findings, and a final evaluation report for the National Test Beds Team. At a local level, they were also asked to share emerging results and progress updates as part of evaluation and governance meetings. This communication was key in managing stakeholder expectations and helping to address emerging challenges, for example, difficulties with patient recruitment and data access.

Figure 4 – Example dashboard of key evaluation information. The information available will be subject to the needs and parameters of individual evaluations

Impact evaluation dashboard

Intervention context

- Aims and objectives
- Target patient cohort
- Geographical scope
- Time period of implementation
- Previous evidence of performance/effectiveness/value for money

Evaluation context

- Aims and objectives
- Key indicators of outcomes
- Method and data used
- Time period of evaluation
- Limitations in the approach

Costs

All costs of combinatorial intervention relative to status quo

- Pilot programme running costs (% incurred if rolled out)
- Capital costs (tech + equipment) total and per patient unit (or other unit)
- Operational costs: total and per patient (or other unit)
- Non-financial costs: management time

Patients/users and recruitment

- X patients recruited across Y sites in Z groups
- Overview of demographic and health status
- Patient engagement or co-production

Impacts

Metric	Result	Comment to interpret results
Benefit-cost ratio	£3.67: £1	Monetised impacts only – excludes XYZ
Cost per QALY	£14,000	Well below the NICE threshold of £30,000/QALY
Cashable savings to the NHS/year	£x, 000	Cashability assumption X
£ cost change for acutes	%	Cashability assumption X
£ cost change for community services	£	
Change in acute A&E admissions and £ savings	X or £X	
Change in primary care use and £ savings	X or £X	
Reported change in wellbeing	Score changes	Various measures, e.g. EQ5D, WEMWBS, GHQ-12, ONS-4

- Key drivers of costs…
- Key drivers of benefits…
- Conditions under which benefits (avoided admissions) were higher are…
- Conditions which constrained benefits being realised were…
- Changes in patient experience were…
- Interdependencies of this intervention with others are…
- Impacts on wider services were…
- Challenges for patients were…
- Further investigation/research needed on…issues
- Potential for results to be replicated if rolled-out
- Potential for economies of scale

Transparency of the final report

Reflections from Wave 1 of the Test Beds programme

Each local evaluation team was tasked with producing a final report which brought together research findings, learning and recommendations from their Test Bed. Although the reports were most likely to be read by a technical audience, and include technical detail about the methods and results used, the National Test Beds Team and National Evaluation Partners requested that the main body of Test Beds' final report should be accessible to everyone. Local evaluators, and programme directors, were asked to present their information using the headings outlined in Figure 5 to help ensure that the document had a clear structure and narrative and addressed the central evaluation questions posed as part of the programme.

Figure 5: example of the final report structure

Section	Overview
Executive summary	Brief overview of the intervention and headline findings for the process and impact/economic evaluations.
Health and care challenge being addressed by the Test Bed	The case for the particular Test Bed intervention.
Overview of the Test Bed intervention	What was implemented, who was involved, intended outcomes, timings etc. logic model and description of theory of change.
Process evaluation	
Aims of the process evaluation	Purpose of the process evaluation and questions addressed in the analysis.
Methods used to collect data and the level / type of role of who has been interviewed	How the evidence was collected and from whom – note that the voice of innovators, clinicians, patients, voluntary sector and commissioners are important (among others as appropriate).
Key findings and interpretation	This must include the evidence to address each of the evaluation questions.
Economic / impact evaluation	
Aims of the impacts / economic evaluation	Purpose of the evaluation and questions addressed in the analysis.
Metrics of interest and data collected	Description of metrics explored, data collection, sample sizes, analytical methods, limitations.
Overview of evaluation methods used	Description of the evaluation analysis carried out (such as diff-in-diff; before vs after etc) and approach for estimating costs.
Key findings and interpretation	Findings in relation to each of the evaluation questions
Conclusions and implications for scale and spread	Insights to inform future investment decisions
Annexes	Technical detail


Section 05

Learning and dissemination

Lessons learned and forward look

The Test Beds programme had several ambitions which were: informing the national evidence base regarding the use of combinatorial innovations and informing commissioning intentions and shaping national practice. Each of these rests on the ability to generate, capture and share learning as much as possible in a timely way so that Test Beds could continually learn and adapt. Wave 1 Test Beds and the National Test Beds Team were therefore proactive in developing and implementing inclusive learning and dissemination activities over the course of the programme.

Reflections from Wave 1 of the Test Beds programme

National and local activity regarding learning and dissemination:

National activity – over the course of Wave 1, the National Test Beds Team and National Evaluation Partner hosted and facilitated 'within programme' dissemination events. These allowed Test Beds to come together to share learning from their evaluation experiences and, where appropriate, to take a group approach to problem solving. These events were quarterly and involved workshops designed to elicit learning; share stories; provide mutual support across the sites; and to facilitate collaborative problem solving. Participants were from across the Test Beds programme including evaluators, clinical leads, service leads, innovators and Programme Directors. These events were helpful in fostering shared learning across sites and sharing insights into the practical challenges faced by Test Bed evaluators and implementers with the National Test Beds Team. Learning was captured, summarised and circulated to all of the Test Beds teams for reference.

Throughout the duration of Wave 1, dissemination events were also held with external parties. The purpose of these events was to share early learning about the practical challenges and learning from evaluating such complex combinatorial innovations. For example, the National Test Beds Team and National Evaluation Partner presented at the King's Fund Digital Health Congress in 2018. The further purpose of such engagement and dissemination was to seek feedback from a diverse audience about what we had learned in terms of evaluation and the extent to which this resonated more widely with other programmes, and hence what could be useful for future programme directors to know if they are setting up similar combinatorial innovations. For example, we hosted a 'Pop Up University' at the Health and Innovation Expo in 2017.

Local activity – Test Beds were expected to build learning and dissemination of research findings into their project plans. For example, prior to the end of the project the Lancashire and Cumbria Alliance Test Bed submitted and had an article published in an academic journal which shared their evaluation protocol and explored the strengths and limitations of their study. In Spring 2018, several Test Beds also submitted abstracts for academic conferences to present the research findings from their final evaluation reports and one site planned an event to launch their local evaluation with local stakeholders such as commissioners. Working in partnership with their local academic and health science network (AHSN) all Test Beds are also expected to develop plans for the promotion of the findings from their evaluation reports and to consider the potential spread and scale of successful innovations.

Forward look regarding national and local activity regarding learning and dissemination:

National activity – To mark the end of Wave 1 and to support the national dissemination of findings and learning from the programme, the National Test Beds Team and National Evaluation Partner are preparing a national evaluation synthesis report which is due to be published in Winter 2018. The document will share economic and impact evaluation data from the seven Wave 1 Test Bed sites as well as process evaluation insights to inform:

- Commissioning decisions about future combinatorial innovations.
- Policy making in terms of where policy attention can best be focused to have greatest impact in the successful delivery of combinatorial innovations.
- A shared understanding about how to build effective NHS – innovator partnerships.

The document will share learning with respect to different types of interventions (for example, predictive analytics, self-management and remote monitoring). It will also explore key themes across all Test Beds including learning around the design and set-up, implementation of interventions, benefits delivered, costs and cost effectiveness; and offer observations about how to evaluate these complex interventions.

The National Test Beds Team and National Evaluation Partner also plan to share the findings and learning contained within the document at both national events and in meetings with policy makers and colleagues across NHS England, the Office for Life Sciences and other government departments. The learning from Wave 1 will be used to inform future waves of the programme.

Learning recommendations

Based on Wave 1 Test Beds experience, it may be helpful for a future wave of Test Beds to:

- Develop learning and dissemination plans from the start of the programme. This is to ensure opportunities are identified early on and continued learning and be facilitated.
- In line with the learning dissemination plans.
- Implement the plans, using a range of learning and dissemination activities and means of engaging others as appropriate.
- Ensure frontline staff, patients and public are among those with whom learning is shared.
- Take stock of learning and embed it in future action.
- Identify key points of learning from the evaluation.
- Identify practical ways to put learning into action for future interventions.
- Translate evaluation findings for policy makers to use.
- Set-up a framework to continue to monitor and evaluate on-going interventions.

Wave 1 Test Beds learning recommendations checklist

Building on the reflections and experiences of Wave 1 Test Beds, a future wave of Test Beds may find it helpful to follow the following steps when conducting evaluation work.

Preparing to commission an evaluation

Being clear what you want:

- ✓ Bring the right stakeholders together and build a collaborative way of working.
- ✓ Clarify the exact scope of the intervention to be evaluated.
- ✓ Decide what learning is required from the evaluation and articulate specific questions to focus on.
- ✓ Decide on the budget and timeframe for the evaluation.

Set things up to maximise continual learning:

- ✓ Identify how and when to improve implementation or share learning.

Ensuring appropriate evaluation governance:

- ✓ Identify the formal reporting required from the evaluators.
- ✓ Map out the advisory groups and quality assurance processes to be set up.
- ✓ Clarify roles, responsibilities and accountabilities of evaluators and others in the evaluation governance structure.

Commissioning an evaluation

Ensuring a fair and open procurement process:

- ✓ Engage the procurement team early and let them guide each step of the process.
- ✓ Establish the skills and experience needed from the evaluation team; what they need to do and when.
- ✓ Prepare a clear and comprehensive ITT and seek feedback on this from relevant stakeholders.

Appointing the evaluation team:

- ✓ Convene an appropriately objective panel of interviewers to select the evaluation team.
- ✓ Agree the questions to be asked at the interviews.
- ✓ Make a jointly agreed decision about who should be appointed as the evaluation team.
- ✓ Keep engaged with the procurement team to ensure all contractual matters are in hand.

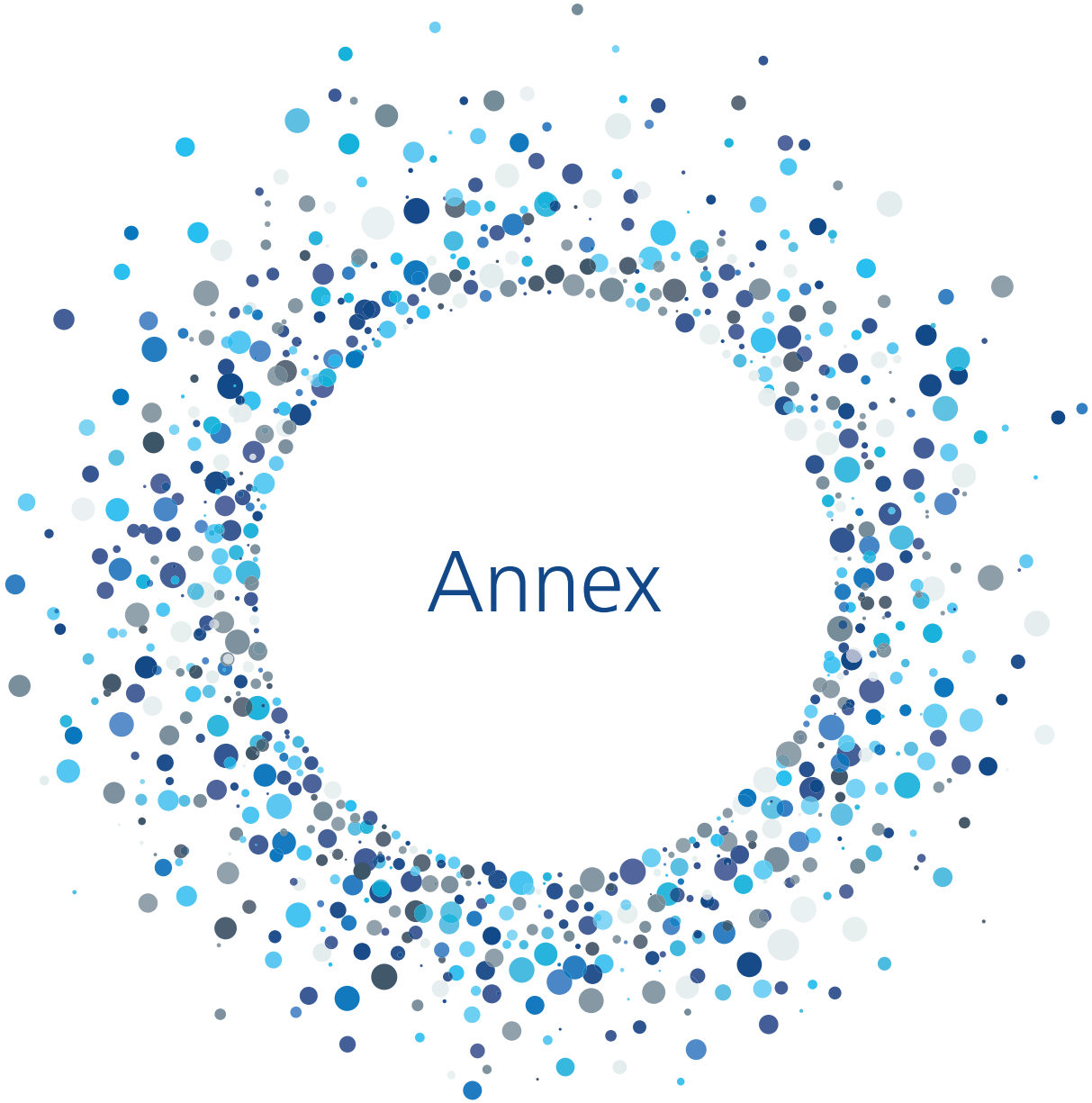
Being an intelligent customer for a robust evaluation

Agreeing the evaluation approach:

- ✓ Bring together the relevant stakeholders to work collaboratively to discuss and agree the evaluation approach.
- ✓ Ensure a clear evaluation framework (logic model) is developed.
- ✓ Be a critical friend to the evaluation team to ensure the evaluation approach is well-justified, appropriate for the particular intervention and proportionate.
- ✓ Provide challenge around data collection and approvals required so that the evaluation is feasible and plans are defined for data collection.
- ✓ Ensure the evaluation team prepares a robust and clear evaluation protocol that is shared and agreed with key stakeholders.
- ✓ Ensure the evaluation team seeks advice from the HRA about ethics clearance and follows due processes to secure relevant clearances.

Being an intelligent customer:

- ✓ Challenge the implementers and evaluators on recruitment targets, progress towards those; and actions to mitigate under-recruitment.
- ✓ Actively ensure communication flows across evaluators, implementers and key stakeholders are working effectively.
- ✓ Provide challenge to ensure results are interpreted appropriately and transparently, recognising uncertainties.
- ✓ Ensure the final report is clear, accessible and robust.


Annex

Annex

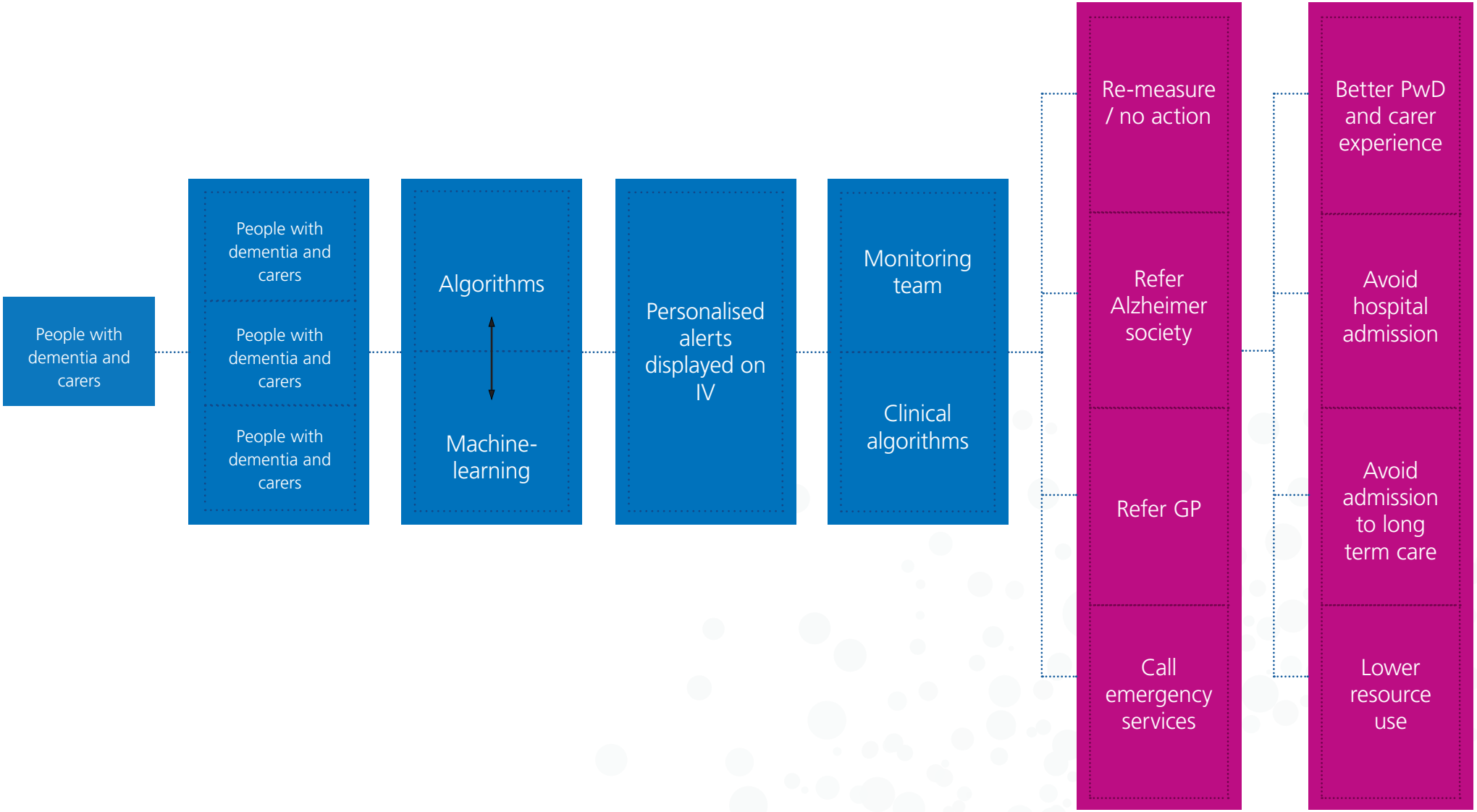
1.1 Steps involved in planning an evaluation – Magenta Guide

Defining the policy objectives and intended outcomes	What is the programme logic or theory about how inputs lead to outputs, outcomes and impacts, in the particular policy context?
Defining the audience for the evaluation	Who will be the main users of the findings and how will they be engaged?
Identifying the evaluation objectives and research questions	What do policy makers need to know about what difference the programme made, and/or how it was delivered? How broad is the scope of the evaluation?
Selecting the evaluation approach	Is an impact, process or combined evaluation required? Is an economic evaluation required? How extensive is the evaluation likely to be? What level of robustness is required? Can proportionate steps be taken to increase the potential for good evaluation? What adjustments to policy implementation might improve evaluation feasibility and still be consistent with overall policy objectives?
Identifying the data requirements	What data are required? What is already being collected / available? What additional data needs to be collected? If the evaluation is assessing impact, at what point in time should the impact be measured? Who will be responsible for data collection and what processes need to be set up? What data transfer and data security considerations are there?
Identifying the necessary resources and governance arrangements	How large scale / high profile is the policy, and what is a proportionate level of resource for the evaluation? What is the best governance structure to have in place? What budget is to be used for the evaluation and is this compatible with the evaluation requirements? Has sufficient allowance been built in? Who will be the project owner, provide analytical support, be on the steering group? What will the quality assurance processes be?

Conducting the evaluation	Will the evaluation be externally commissioned or conducted in-house? Who will be responsible for specification development, tendering, project management and quality assurance? When does any primary data collection need to take place? Is piloting or cognitive testing of research instruments required? When will the evaluation start and end?
Using and disseminating the evaluation findings	What will the findings be used for, and what decisions will they feed into? How will the findings be shared and disseminated? How will findings feed back into the ROAMEF cycle?

Source: H M Treasury, (2011), The Magenta Book: Guidance for evaluation, HM Treasury: London

1.2 An example of a logic model used by a Test Bed in Wave 1: Logic model for TIHM for dementia



Process evaluation

Process evaluations can use a variety of qualitative and quantitative techniques to explore how a policy was implemented describing the actual processes employed, often with assessments of the effectiveness from individuals involved or affected by the policy implementation

Empirical impact evaluation

Empirical impact evaluations use quantitative data to test whether a policy was associated with any significant changes in outcomes of interest. Various approaches are available which differ in their ability to control other factors which might also affect those outcomes (the counterfactual, either directly measured or imputed) and hence in the confidence it is possible to place in the results.

Economic evaluation

Economic evaluation involves calculating the economic costs associated with a policy, and translating its estimated impacts into economic terms to provide a cost-benefit analysis. (When only a costing exercise is undertaken, the result is a cost-effectiveness analysis.) Economic evaluations will often make use of existing evidence and assumptions to facilitate the translation of inputs and actual measured outcomes into economic measures, making them akin to theory-based evaluations (see below). The HM Treasury Green Book provides detailed guidance on economic evaluation and cost-benefit analysis.

Theory-based evaluation

Theory-based evaluation approaches involve understanding, systematically testing and refining the assumed connection (i.e. the theory) between an intervention and the anticipated impacts. These connections can be explored using a wide range of research methods (both qualitative and quantitative), including those used in empirical impact evaluation

Meta-evaluation and meta-analysis

Meta-evaluations (covered in more detail in Chapter 6) can use quantitative or qualitative techniques to bring together a number of related evaluations to derive an overview or summary conclusion from their results.

Simulation modelling

Simulation modelling is one way in which the results of different evaluations of separate parts of the impact pathway or logic of an intervention can be combined and requires that the evidence relating to the different links in the logic model are expressed in quantitative terms (e.g. effect sizes).

3.0 Types of resources employed in evaluation – Magenta Guide

Resource type	Description
Financial resources	A substantial part of the costs of an evaluation may be incurred after the policy has been implemented. Therefore, it is important to think about the financial resources required for the evaluation whilst planning the policy budget. Cost will be substantially lower if data can be used which already exist and/or are being collected through monitoring activities. Data collection exercises might need to be funded if the policy is novel or targeting unusual or hard-to-measure outcomes.
Management resources	Both internal and external evaluations will often require a dedicated project manager (with the specialist technical expertise to assure quality) who is responsible for: commissioning (for external evaluations); day-to-day management; advising the evaluation contractors and reacting to issues that develop. The level of input required will be greatest at key points (in particular, the design and commissioning stage), but this will be an ongoing resource requirement and should not be underestimated.
Analytical support	Due to the multi-disciplinary nature of many evaluations, it is important to consider the range of internal analytical specialists (such as social researchers, economists, statisticians, operational researchers, or occupational psychologists) who might need to be called upon for advice and to help design the evaluation approach and outputs. They can also advise on the effect of policy design on the feasibility of undertaking different types of evaluation. This can help ensure that the evaluation design will provide evidence to answer the research questions, and that, if necessary, appropriately skilled contractors are commissioned. Analytical input can also be useful in the steering of the project and in the quality assurance of outputs.
Delivery bodies	A successful evaluation will often depend crucially on the early and continued engagement and cooperation of the organisations and individuals involved in delivering the policy. It will be important to communicate what the evaluation seeks to address, what input will be required from them, and how they might benefit from the findings.
Wider stakeholders	The evaluation may also involve other stakeholders – for example, people and organisations directly or indirectly affected by the programme. The level of involvement and method of engagement will be specific to the policy and stakeholders in question, but may include inviting them onto a steering group, informing them about the evaluation, or including them as participants in the research.
Peer review	In order to ensure quality it may be necessary to have aspects of the evaluation peer reviewed. This is a requirement in some central government departments. Peer review might include the methodology, the research tools, and any outputs including interim and final reports.

Source: H M Treasury, (2011), The Magenta Book: Guidance for evaluation, HM Treasury: London

3.1 Factors affecting appropriate resourcing of an evaluation – Magenta Guide

Factor	Explanation
Innovation and risk	High risk policies are likely to require robust evidence to understand both how they are working in practice and whether they are having the predicted impacts. In those cases where the innovative initiatives might offer 'low cost solutions' evaluation resources might be 'disproportionately' high but are still needed to demonstrate the scale of the returns on the policy investment.
Scale, value and profile	Large scale, high-profile, or innovative policies or policies that are expected to have high impact are likely to require thorough, robust evaluation to help build the evidence base on what works, meet accountability requirements, assess returns on investment and demonstrate that public money is well spent.
Pilots	Pilot or demonstration projects, or policies where there is a prospect of repetition or wider roll out, require evaluation to inform future activities.
Generalisability	If it is likely that the findings will have a much wider relevance than the policy being evaluated, more resource may need to be allocated to ensure that the results can be generalised with confidence.
Influence	If the evaluation is capable of providing information which can have a large influence on future policy (for example, it can report at a strategic time point and/or meet a key evidence gap) more resource is likely to be justified.
Variability of impact	The effects of policies with highly uncertain outcomes or with significant behavioural effects are likely to be more difficult to isolate, and there is likely to be a greater case for conducting a more extensive evaluation.
Evidence base	Where the existing evidence base is poor or under-researched an evaluation is likely to require more resources in order to fill the gaps.

Source: H M Treasury, (2011), The Magenta Book: Guidance for evaluation, HM Treasury: London

4.0 Issues affecting the choice of evaluation approach – Magenta Guide

Evaluation objectives and research questions

The overall objectives of the evaluation and the specific research questions it needs to answer are important factors in deciding which evaluation approach(es) to use and should be developed from the logic model. General research questions which are not overly specific to the intervention in question might be answerable via a qualitative review (or more formal analysis) of the existing literature. Questions which are more specific to the intervention will involve one of the other evaluation types listed in Box 5.B. Questions relating to the wider or ultimate objectives of an intervention will generally require some form of impact evaluation – possibly as part of a theory-based evaluation approach if the associated impact pathways are very extended or complex. Questions relating to detailed aspects of the workings of the policy will generally imply some form of process evaluation (although a combined impact evaluation might be warranted if more definitive answers about effectiveness are required).

Complexity of the logic model and importance of confounding factors

Where the logic model is particularly complex, restricting the scope of the evaluation to consider shorter, simpler "links" in the logic chain can increase the ability of process evaluations to provide good evaluation evidence. However, if significant confounding factors remain, a robust impact evaluation with suitable controls might be necessary to generate reliable findings. The feasibility of this might depend on data availability (for quasiexperimental approaches) and time and resources (for approaches needing dedicated data collection). Detailed evaluation of changes in very complex systems (especially those with a significant geographical component) might only be possible through theory-based evaluation or simulation modelling.

Availability and reliability of existing evidence

Large amounts of strong existing evidence increase the relevance of review based methodologies, facilitate greater use of simulation models, and enable evaluations to be simplified to focus more closely on those specific questions which the current evidence base leaves unanswered.

Existing data sources and measurability of outcomes

If there is already a wide range of good quality data sources covering outcomes of interest, the feasibility of undertaking robust impact evaluations (sometimes to relatively short timescales) is greatly increased. Outcomes which are difficult to measure require either dedicated data collection (e.g. through surveys) or a way of estimating them from changes in intermediate indicators. The former implies a more resource- and time-intensive study, as does a lack of existing data (which might be the case particularly when the focus of the evaluation is the specifics of a very localised intervention). The latter might be addressed through a simulation model, subject to existing data availability.

Time and resource availability

In most cases, process evaluations (including action research and case studies) will require a formal commission and a dedicated research team, often externally contracted. This can imply a considerable time and resource commitment. Impact evaluations requiring specific data collection and outcome measurement can similarly involve heavy resource commitment and long project durations. Impact evaluations which are able to use existing datasets can provide rigorous results in relatively short timescales but this same reliance on existing data can restrict the questions they can attempt to answer and, in some cases, the ability to confidently attribute the impacts to the intervention. Simulation models can also sometimes be undertaken relatively quickly but this depends on a range of assumptions being made to limit their scope.

Empirical impact evaluation issues

The two principal strengths of empirical impact evaluation approaches are that they can isolate the effect of an intervention from the possible multitude of factors which might have an influence on the outcome of interest; and in this way, they can provide a rigorous test of whether the intervention has an effect or not. However, these strengths can come at a cost. That is that the approaches are often less able than other approaches to explain exactly why any difference occurred (or not), or how it varied across circumstances. Much of this can (and should) be overcome by using a mixed design, whereby process and impact evaluations complement each other, and the process evaluation can help to explain the impact evaluation findings. In other cases based on statistical regression analysis the relationship between the intervention and the outcome of interest might be so complex that the evaluation will only be able to say whether the intervention had an effect, not what aspects of it, how or why. Some 'procedural' explanation might be possible, but only if the scope of the evaluation is restricted to simpler relationships, for instance, between the intervention and some intermediate outcome rather than the ultimate objective of the intervention (e.g. the impact of the intervention on the take up of training, rather than the impact on employment and wages).

Source: H M Treasury, (2011), The Magenta Book: Guidance for evaluation, HM Treasury: London

5.0 Interpretation of Results – an explanatory list

Strength of findings on process and drawing on qualitative methods:

- What was the purposive sampling strategy used to select evaluation participants? Were all key groups included in the evaluation? Were any key groups missing and how does this limit the evaluation findings?
- Were topic guides or other fieldwork instruments well-designed, in order to support the systematic collection of qualitative data, consistent from one setting and researcher to the next?
- What method of analysis was used and why?

Strength of findings on impacts:

- Given the number of patients recruited to the intervention, was this sufficient for the findings to be robust (statistically significant or of a sufficient power)?
- What are the key assumptions and how do these affect the results?
- Have sensitivity tests been carried out to explore how the results change under different assumptions?
- What are the limitations of the data used or analysis carried out and what do these mean for the findings?
- Have costs (inputs) been robustly estimated?
- How are different patient groups affected by the intervention and to what extent are some affected more than others, and why?

Headline findings:

- Are there clear and robust findings?
- What are the key takeaway messages and results from the evaluation?
- How cost effective was the intervention?
- How can NHS-commercial partnerships be facilitated to work well?

Unintended consequences:

- Has the evaluation appropriately explored the extent to which unintended impacts have occurred? What are they and why have they come about?

Scaling up:

- What can we take away from the analysis to inform scaling-up decisions?
- What would need to be different if the intervention were scaled up?
- What are the conditions under which the intervention is likely to be most effective and why?

Source: H M Treasury, (2011), The Magenta Book: Guidance for evaluation, HM Treasury: London