

NHS Cancer Programme Innovation Open Call: An Introductory Guide to Evaluation

NHS England



Contents

1 Purpose.....	2
2 Introduction to evaluation	2
2.1 What is evaluation?.....	2
2.2 Why undertake evaluation?.....	2
3 Distinguishing between monitoring and evaluation.....	2
3.1 Monitoring	3
3.2 Evaluation	3
4 Developing a Theory of Change for evaluation.....	3
5 Approaches to evaluation	4
5.1 Formative evaluation.....	4
5.1.1 Process evaluation	4
5.2 Summative Evaluation	6
5.2.1 Impact evaluation	6
5.2.2 Economic Evaluation	7
5.3 Understanding impact	9
6 Data collection.....	10
6.1 Routinely collected data.....	11
6.2 Bespoke quantitative data	11
6.3 Bespoke qualitative data.....	11
7 Mapping evaluation approaches to a logic model	11
8 Additional resources.....	19

1 Purpose

This guide sets out:

- An introduction to evaluation and why it is important,
- The key methodological approaches to evaluation, and
- Data collection strategies

The guide is available to the NHS Cancer Programme's Innovation Open Call applicants to support them to develop their evaluation as part of their application. Every innovation is required to complete an independent evaluation to determine the impact of their innovation and develop the evidence based required to support large-scale implementation.

2 Introduction to evaluation

2.1 What is evaluation?

Evaluation examines the implementation and impacts of a policy or intervention to assess whether the anticipated effects, costs and benefits were in fact realised. Evaluation findings can identify “what works”, what doesn't and why, highlight good practice, identify unintended consequences or unanticipated results, and demonstrate value for money. Evidence generated through evaluation should be fed back into the programme or policy cycle to improve future decision-making.

Evaluation goes beyond reporting a single statistic or data item and helps us to understand whether a situation is better or worse than before, how it might compare to other areas, or how it compares to an alternative approach.

2.2 Why undertake evaluation?

All policies, programmes and projects should undertake evaluation. By not undertaking evaluation, we will not know whether a policy or intervention was effective, or worse, resulted in overall perverse, adverse or costly outcomes.

Evaluation is a key enabler to improving cancer services, particularly where the evidence base is less established. Evaluation findings can also indicate where we can make changes to services and interventions, which can lead to better outcomes for both patients and staff, and therefore help commissioners and providers in their decision making and allocation of resources.

Evaluation also contributes valuable knowledge to the evidence base, feeding into future policy development, thus occupying a crucial role in the policy cycle.

3 Distinguishing between monitoring and evaluation

While there is some overlap between monitoring and evaluation there are distinct differences:

3.1 Monitoring

Monitoring can be defined as the formal reporting and evidencing to ensure that inputs and outputs are successfully delivered, and implementation milestones met. Analysis of monitoring data against planned targets will help to identify whether a specific programme is being implemented as expected and where further action may be required.

3.2 Evaluation

Evaluation is the assessment of a programme's effectiveness and efficiency during and after implementation. It seeks to measure the effect of a policy or intervention on planned outcomes and assess whether the anticipated benefits have been realised, how this was achieved, or if not, why not.

4 Developing a Theory of Change for evaluation

A robust evaluation helps to determine which policies and programmes are most effective in which circumstances, what difference they have made and whether their overall benefits justify the costs. Developing a robust evaluation is dependent on having a thorough understanding of what it is the programme or policy is trying to achieve and how. This may be referred to as the logic or theory behind the programme of work, which sets out the problem you are trying to address (the rationale or situation), what you want to achieve (the goal or impact) and how you are going to achieve the desired impact through a series of activities and expected outputs and outcomes.

Developing a Theory of Change can help you to do this. A Theory of Change is a description of how and why a desired change is expected to happen in a particular context and is a useful approach for programme management and evaluation purposes. Specifically, a Theory of Change explains the process of change by outlining causal links in an initiative – setting out what you plan to do and the outcomes you expect to see. These links are then explained by the assumptions underpinning why a specific activity is expected to lead to a predicted outcome and how a given outcome will lead to longer-term impacts.

A Logic Model (see Figure 1 below) is one way to articulate your Theory of Change. A logic model is a graphic display of the relationship between a programme's resources, activities and intended results, which identifies the programme's underlying theory and assumptions and is a useful evaluation and programme management tool.

Your Theory of Change or Logic Model will help you to define the questions that you need to address, which will inform the design of your evaluation and required resource. It can then help you to determine if the intervention or programme is being

implemented as planned and if the outcomes you observe are the ones that you expected.

Figure 1: Simple Logic Model Showing the Different Types of Evaluation



5 Approaches to evaluation

The key approaches to evaluation set out here can be categorised as either formative or summative. Briefly, process evaluation is a formative evaluation approach, while impact and economic evaluation are summative evaluation approaches. The next section introduces these principal approaches to evaluation in further detail.

5.1 Formative evaluation

The primary purpose of formative evaluation is to gather information and provide ongoing feedback that can be used to improve or strengthen the implementation of a programme. Formative evaluation takes place during the development of a project, intervention, or service redesign, to ensure the intervention is feasible and appropriate as it is being implemented and before it is fully implemented.

5.1.1 Process evaluation

How was the programme or policy delivered?

Process evaluation seeks to determine whether an intervention has been implemented well and is operating as planned. In general, process-related questions are intentionally descriptive, and as a result, process evaluations can employ a wide range of data collection and analysis techniques, covering multiple topics and participants, tailored to the processes specific to the policy in question.

Process evaluations will often include the collection of qualitative and quantitative data from different stakeholders, using interviews and surveys. These might cover subjective issues (e.g., perceptions of how well a policy has operated) or objective aspects (e.g., factual details of how a policy has operated). They might also be used to collect organisational information (e.g., how much time was spent on specific activities), although “administrative” sources (e.g., timesheets and personnel data) might be more reliable, if available.

Although essentially descriptive, these types of information can be vital to measuring the inputs of an intervention (which might include simple financial budgets but also staff and other resources) as well as the outcomes (e.g., surveys might be used to

measure aspects of participants' [quality of life](#)). This illustrates the practical link between process and impact evaluations, which often implies a need to consider the two together.

A useful guide on how to carry out a process evaluation in complex systems can be found [here](#).

Figure 2: Process Evaluation [Example](#)

Multi-Disciplinary Team (MDT) streamlining project: process evaluation

In 2018, the NHS Cancer Programme carried out a process evaluation of the Multi-Disciplinary Team (MDT) streamlining project.

Policy context

Multi-disciplinary team meetings (MDTMs) are subject to a range of pressures highlighted in academic literature, reports (CRUK, 2016; Taskforce Report, 2015) and consultations. More people are being diagnosed with cancer, and surviving for longer, meaning more patients are discussed at MDTMs. The MDTM has not evolved to cope with this demand. Almost 50% of patients are discussed for less than 2 minutes and meetings often last for hours (CRUK, 2016). This is unsustainable and inefficient.

In 2018, a number of MDTs tested an approach to streamlining patients with the objective of shifting time from straightforward cases to more complex ones, and potentially reducing time spent overall.

Evaluation

A process evaluation was undertaken to understand the different approaches pilot sites took to implementation during the testing period. Specifically, the evaluation sought to understand the broader implications of this practice change for MDTs, Providers, and other stakeholders:

- *What impact the introduction of the Standards of Care (SoC) had on the workload of MDTM staff?*
- *How the SoC was implemented, including triage processes; whether there were barriers and enablers to implementation; and how these were identified and overcome?*

The NHS Cancer Programme conducted a number of in-depth interviews with key stakeholders across participating sites to gain an in-depth understanding of local approaches to implementation and any challenges and unintended consequences. Evidence generated through this evaluation informed the development of national policy while being proportionate to the scale of the programme.

5.2 Summative Evaluation

Summative evaluations are intended to show whether the programme has achieved its intended outcomes (i.e., intended effects on individuals, organizations, or communities) and to indicate the ultimate value of the programme at its conclusion. Summative evaluations seek to determine whether the programme should be continued or ended.

Summative evaluations include **impact (or outcome) evaluations** and **economic evaluations**.

5.2.1 Impact evaluation

What difference did the policy make?

An impact evaluation seeks to determine to what extent the programme is achieving the expected outcomes and longer-term impacts. It seeks to determine longer-term changes that have occurred from the programme. These impacts are the net effects (including unintended consequences) on programme stakeholders (including the target population).

Answering the question of 'what difference a policy has made' involves a focus on the outcomes and longer-term impacts of the policy. Impact evaluations will be measured by outcomes that demonstrate impact. The outcomes therefore are measurable achievements which are aligned to the objectives of the policy.

There may be [challenges](#) when measuring the impact of a programme. Below lists some of the challenges that might be encountered:

- Traditional approaches can underestimate impact and value
- Focus on spend can lead to missing some aspects of value
- System change has multiple players
- System change takes time
- Capturing and using evidence well takes expertise and confidence

Figure 3: Impact Evaluation Example

Be Clear on Cancer awareness campaign: Impact Evaluation	
Evaluation Questions	Metric
Are people seeing the campaign and is it raising awareness of the signs and symptoms?	Cancer and Campaign awareness
Are we seeing more people going to their GP with the symptoms promoted by the campaign, and is there a shift in the profile of patients presenting?	GP attendance
Are we seeing more people referred urgently for suspected cancer, and is there any shift in the profile of these patients?	Urgent referrals for suspected cancer
Of those referred urgently for suspected cancer how many turn out to have that cancer?	Conversion rates
Are we seeing an increase in diagnostic investigation activity, or the length of time patients are waiting for tests?	Impact on investigations
Are we seeing an increase in the numbers of patients diagnosed with cancer, and/or a shift towards earlier stage disease?	Cancer incidence and staging

Evaluation findings are presented as individual metric summaries and subsequently as final evaluation reports. [These reports and methodological details can be downloaded from the NCRAS website.](#) Earlier evaluation reports are available from [CRUK](#).

5.2.2 Economic Evaluation

Did the benefits justify the costs?

Economic evaluation is the process of identifying, measuring and valuing programme inputs and outcomes of two alternative activities, and the subsequent comparative analysis of these. There are 3 main types of economic evaluation:

- [Cost-utility analysis](#)
- [Cost-effectiveness analysis](#)
- [Cost-benefit analysis](#)

Economic approaches value inputs and outcomes in quite particular ways, and it is crucial that the needs of any economic evaluation are considered at the design stage. Otherwise, it is very likely that the evaluation will generate information which, although

maybe highly interesting and valid, is not compatible with a cost-benefit framework, making it very difficult to undertake an economic evaluation.

A health economic evaluation will measure two parameters (i) the costs of delivery of a given health care intervention, and (ii) the benefits of the intervention (or outcomes).

Some elements of a health economic evaluation can be particularly hard to attribute to a given intervention or even measure such as quality of life (QoL). A number of tools and resources are available to support the measuring of QoL, including [EQ5D](#), [EORTC](#) and [FACT](#).

Figure 4: Outcome Evaluation Example

Replacing guaiac faecal occult blood test (gFOBT) with faecal immunochemical test (FIT) in the England bowel screening programme: economic evaluation

The Faecal Immunochemical Test (FIT), an improved colorectal cancer screening test, has been recently introduced in England, Scotland and for a small number of people in Wales. This test works similarly to the current standard guaiac-based faecal occult blood tests (gFOBT); however, pilot studies suggest FIT to be more practical as it is specific to human haemoglobin and is quantifiable. In addition, studies have found that it detects twice as many cancers, is more accurate, has adjusted sensitivity and is found to be more favourable to patients resulting in increased uptake. To provide information on the longer health and economic consequences and help in choosing one test over the other for screening, a health economic analysis is required.

Economic evaluation:

In July 2015 a cost-effective evaluation analysis of FIT versus gFOBT for colorectal cancer screening for a UK population aged 60 – 75 years was published. This analysis aimed to estimate the cost-utility of screening using FIT compared with gFOBT in the National Health Service Bowel Cancer Screening Programme (NHS BCSP) in England. A mathematical model was constructed to estimate the difference in incremental costs (cost of FIT screening minus cost of gFOBT screening) and health quality outcomes of screening using FIT and gFOBT kits. Data on screening uptake, detection, adverse event, motility and costs was collected from the BCSP and other National sources for the analysis.

Outcome:

This work informed decision makers on predicted resource cost, costs and quality of life outcomes when using FIT kits for screening in NHS BCSP. The result suggested that FIT is more effective and less costly compared to gFOBT at all thresholds – it achieved greater health gains and with additional cost savings in cancer management. Further details of the study and the results can be found [here](#).

As a result of this study and many others the FIT kit has now been introduced and will gradually replace gFOBT in England as of June 2019. The FIT kit is now sent out as a standard test with all invitations to people eligible for bowel screening.

5.3 Understanding impact

The key characteristic of a good impact evaluation is that it recognises that most outcomes are affected by a range of factors, not just the policy. To test the extent to which the policy was responsible for the change, it is necessary to estimate what would have happened in the absence of the policy. This is known as the counterfactual.

Establishing the counterfactual is not easy since, by definition, it cannot be observed – it is what would have happened if the policy had not gone ahead. A strong evaluation is one which is successful in isolating the effect of the policy from all other potential influences, thereby producing a good estimate of the counterfactual. An evaluation might also be able to explain how different aspects of the policy contributed to the impact.

A robust counterfactual is constructed from a similar population living in similar circumstances, but where the intervention or service change has not been introduced. However, in practice, determining the counterfactual in complex health systems can be challenging. There are 3 levels at which the counterfactual can be achieved:

- Experimental options which are achieved through developing a counterfactual using a control group. Participants are randomly assigned to either receive the intervention or to be in a control group (Randomised Control Trial). This is often referred to as the gold standard for measuring impact. Please note that randomised control trials are not eligible for the NHS Cancer Programme's Innovation Open Call evaluations.
- Quasi-experimental options which are achieved through the development of a counterfactual using a comparison group which has not been created by randomisation. These aim to establish a cause-and-effect relationship between an independent and dependent variable but does not rely on random assignment.
- Non-experimental options such as the development of a hypothetical prediction of what would have happened in the absence of the intervention. This can most simply be achieved by measuring the factors of interest before and after the programme or policy has been implemented.

Whether a robust impact evaluation is possible depends on features of the policy or intervention itself, the targeted outcomes, and how well the evaluation is designed. If a robust evaluation is not possible, or the evaluation is poorly designed, the estimated counterfactual will be unreliable, and there will be uncertainty over whether the outcomes would have happened anyway, regardless of the policy. Under these conditions, it will not be possible to say whether the policy was effective or not, and even if policy outcomes appear to move in desirable ways, any claims of policy effectiveness will be unfounded.

More information on establishing a counterfactual can be found [here](#).

Longer term impacts can often be difficult to measure in the time frame of the programme. As such, intermediate measures, or proxy measures such as the stage

that cancer is detected, are often used. Intermediate or proxy measures should be relevant to the hypothesised longer-term impacts.

While measuring impact through the use of a counterfactual is the preferred and most robust approach, it can be challenging to establish a strong counterfactual in complex, real world, health care systems. Another way to test whether the outcomes you observe is due to the intervention you have put in place is to carry out a contribution analysis. Here stakeholders judge the size and value of the contribution of the work that has been carried out to the outcomes and overall goals that have been achieved. Contribution analysis can be particularly useful in complex systems where other factors may influence the outcomes that have been observed.

Contribution analysis is a key methodological approach in the NHS England Sustainable Improvement [Impact Framework](#).

Further information on how to understand and measure impact can be found [here](#).

Finally, it can be tempting when establishing evaluation to jump straight to measuring impact. However, when undertaking a summative evaluation, it is important to also consider process and economic questions in recognition that there are likely to be several factors that may have contributed to observed outcomes. That is, the outcomes you observe could be due to factors other than the intervention or programme that you are testing.

6 Data collection

Evaluation can employ a variety of analytical methods to gather and assess data and information, and the choice of methods employed will depend on a wide range of factors including:

- The availability of the required information and associated lags in routinely collected information,
- Available timescales for primary or bespoke data collection,
- The format the information is collected in,
- Who will be responsible for collecting the information, and
- Important information governance considerations.

Information should be collected to address detailed descriptions of what was involved in the programme, how it was delivered, the difference it made, costs, and the experience of participants and those involved in its delivery. You should ensure that the specific information required is directly related to the questions you are seeking to answer through the evaluation.

Broadly, the information collected for evaluation purposes can be categorised in three groups:

- Routinely collected data
- Bespoke quantitative data
- Bespoke qualitative data

6.1 Routinely collected data

Some of the data you require may already be routinely collected and nationally available. It is important therefore to map required information and metrics against the key questions the evaluation is seeking to address. This data may include qualitative as well as quantitative sources. It is important to note that much of this data may have a time lag between collection and publication.

6.2 Bespoke quantitative data

You may need to collect some data locally, that which is not already routinely available. There are a number of useful online guides that you may find helpful, including resources developed by NHS Improvement: [Measurement for Improvement: an overview](#), [the How-to guide for measurement improvement](#), [Seven steps to measurement improvement](#) and [the Interactive improvement measurement tool](#).

6.3 Bespoke qualitative data

To capture data on staff, patients, and the public's experience of services you may need to collect qualitative data. This information can be collected in a number of ways including through surveys, interviews or focus groups. Qualitative data is a key part to comprehensively understanding and describing the impact of a programme or intervention and can be particularly useful in highlighting any unintended consequences.

7 Mapping evaluation approaches to a logic model

A completed logic model for an intervention to increase cervical screening uptake among Black, Asian and Minority Ethnic (BAME) women is presented in Table 1 below. Corresponding evaluation questions, as mapped to this specific logic model, are presented thereafter in Table 2. This example is provided to illustrate the role of a logic model in informing the evaluation design, and the type, and level, of information required to establish and undertake evaluation.

Table 1: logic model for an intervention to increase cervical screening uptake among BAME women.

Programme objective: to increase cervical screening uptake among BAME women by 5% across the whole Cancer Alliance by year end.

	Formative Evaluation			Summative Evaluation	
	Process Evaluation			Outcome and Impact Evaluation	
Rationale	Inputs	Activities	Outputs	Outcomes	Impact
<p>Cervical cancer is the 14th most common cancer in UK, accounting for 3,200 new cases are diagnosed every year (around 2% of all new cancers diagnosed each year in women), although incidence rates are projected to increase.</p> <p>BAME women are more likely to report that they have never attended a cervical screening test than white women; less likely to report the benefits of screening and less comfortable talking to a GP about cervical screening 99.8% of cases are preventable.</p> <p>While the NHS target is for 80% of women to be tested, uptake nationally is at approximately 72%, with</p>	<p>Cancer Transformation Funding and CCG baseline funding.</p> <p>IT infrastructure & support.</p> <p>Project management resource.</p> <p>Data & analysis.</p> <p>Management and clinical leadership and support.</p> <p>Clinical capacity for onward referrals.</p> <p>Clear protocols and processes established to support screening pathway.</p>	<p>Launch series of targeted campaigns to raise awareness and understanding of cervical cancer and the benefits of cervical cancer screening.</p> <p>Drawing on best practice, develop a suite of resources targeted at BAME women to encourage uptake of cervical screening.</p> <p>Work with GP practices with the highest proportion of BAME registered patients and wider community services to develop and implement a 'Making Every Contact Count' strategy to encourage cervical screening uptake.</p> <p>Invite patients to attend local GP practice for a</p>	<p>Strategy implemented in selected GP practices.</p> <p>Health professionals proactively engage with BAME groups to encourage uptake of cervical screening during all contact opportunities.</p> <p>Wider roll out across whole Cancer Alliance to ensure greatest opportunity to increase cervical screening uptake.</p> <p>Patients with abnormal results are referred on to Specialist Screening Practitioner (SSP) clinic as per established protocol.</p> <p>Patients diagnosed with cancer are put onto a cancer pathway.</p>	<p>Measurable impact on uptake of the cervical cancer screening programme among BAME women of 5%, and reduced variation between BAME and non-BAME populations.</p> <p>More cancers are detected at an earlier stage than they otherwise would have been.</p> <p>Fewer cancers are detected through emergency presentation.</p>	<p>Improved survival among BAME women.</p> <p>More women living with and beyond cancer have a better quality of life.</p>

	Formative Evaluation			Summative Evaluation	
	Process Evaluation			Outcome and Impact Evaluation	
Rationale	Inputs	Activities	Outputs	Outcomes	Impact
<p>marked variation nationally.</p> <p>Increased uptake in screening will lead in more cancers being detected at an earlier stage, which is associated with improved survival, fewer treatment complications and a better quality of life.</p>		<p>cervical smear test and send reminder letters to patients who do not attend.</p> <p>Smear tests are analysed by screening hub, and proceed as per protocol, depending upon test result.</p>			
Example assumptions	There is sufficient local clinical capacity for onward referrals.	This is appropriate promotion of screening programme to encourage patients to participate, which will in turn lead to an increased rate of uptake.	<p>Targeted campaigns lead to an increased uptake of cervical screening among BAME women.</p> <p>Patients referred on to SSP clinic attend appointment.</p>	Patients are diagnosed with cervical cancer earlier than they would have otherwise been.	A greater number of patients are diagnosed with cervical cancer earlier than they would have otherwise been resulting in improved survival and a better quality of life for those patients.

Table 2: specific evaluation questions, by evaluation approach, and corresponding data sources.

Questions	Evaluation approach	Data collection methods	Data sources
<ul style="list-style-type: none"> • Why has this programme been chosen and what evidence is there for the need? • What is the programme aiming to achieve? 	Needs Assessment	<ul style="list-style-type: none"> • Review of pathway models and local evidence/ documentation • Published evidence on the effectiveness of the programme 	
<ul style="list-style-type: none"> • What problem is the programme aiming to address? • What is the programme aiming to achieve? • What approach(es) were used to increase screening uptake? • Who was the target population and why were they selected? • What were the barriers and enablers to implementation? • What did staff and participants feel worked, what didn't and why? • What were participant's experience of the programme? 	Process evaluation	<p>Review of programme documentation</p> <p>Bespoke qualitative data collection - surveys, interviews and/or focus groups of staff and patients to understand what works and what doesn't; and/or of patient experience</p>	
<ul style="list-style-type: none"> • What impact has the programme had on: <ul style="list-style-type: none"> ▪ take-up rates; ▪ detection of cancer (detection rates); ▪ stage of diagnosis; ▪ survival; ▪ service demand and activity locally (including on Cancer Waiting Times); ▪ patient experience and ▪ reducing variation and inequalities (e.g. access, uptake or outcomes)? 	Outcome/Impact evaluation	Routinely collected data	<p>Screening uptake and coverage (CancerStats2; various official statistics)</p> <p>Routes to diagnosis</p> <p>Stage of diagnosis</p> <p>Survival</p>
		Bespoke quantitative data – survey; data collection on activity and participants	

<ul style="list-style-type: none"> • What were the costs of the programme (for example costs of awareness campaign and/or invites/reminders; increased screening costs from higher uptake; bringing forward and potential changes in treatment costs from earlier diagnosis)? • What were the benefits or savings (for example potential changes in treatment costs from earlier diagnosis and improved quality of life (quality adjusted life years (QALYs))?) • Do the benefits/savings from the programme outweigh the costs? 	Economic evaluation	Routinely collected data e.g. unit costs	Average cancer incidence costs per patient for breast and colorectal cancers by stage NHS reference costs Unit costs of health and social care
		Bespoke quantitative data – data collection on activity, (unit costs and activity)	

8 Additional resources

The following is a list of selected readily available resources, designed to support policy, operational, and analytical colleagues to undertake evaluation.

[HM Treasury Green Book](#)

HM Treasury guidance on how to appraise and evaluate policies, projects and programmes.

[Better Evaluation](#)

A range of useful resources and guides to different approaches to evaluation

[NHS England Impact Framework](#)

Sets out a framework to systematically capture the impact of these programmes of work

[Better Care Fund. How to understand and measure impact](#)

Aims to help local areas to understand and measure the impact of their efforts to integrate services across the provision of health and care.