Evaluation: what to consider

Commonly asked questions about how to approach evaluation of quality improvement in health care

The Health Foundation

Contents

Introduction

Evaluation is an essential part of quality improvement and when done well it can help solve problems, inform decision making and build knowledge. While evaluation comes in many shapes and sizes, its key purpose is to help us to develop a deeper understanding of how best to improve health care.

People involved in quality improvement often ask the Health Foundation about how to approach evaluation. Inspired by the most commonly asked questions, this guide is intended to assist those new to evaluation by suggesting methodological and practical considerations and providing resources to support further learning.

We have not intended to produce a prescriptive, step-by-step guide to evaluation as people and organisations will have very diverse evaluation needs. Instead, we aim to stimulate your thinking and support your plans for evaluation.

You can read this booklet from cover to cover, or you might prefer to dip into particular areas of interest. We hope that this will help you think through your approach to evaluation.

Box 1: What is evaluation?

There are many definitions of evaluation, but here are three we like:

- The process of determining the merit, worth or value of something.[*]

- Using systematic, data-based inquiries about whatever is being evaluated.[†]

- A process undertaken for purposes of improvement, decision making, enlightenment, persuasion.[‡]

An evaluation has to be specifically designed to address the questions being asked and the nature of the intervention being evaluated. This means using different methods, working in different settings, with varied populations and data, under specific constraints of time, expertise and resources, both human and financial.

---

[*] Scriven M. *Evaluation thesaurus.* Sage Publications, 1991.

[†] American Evaluation Association (www.eval.org), 2004

[‡] Shadish W, Cook T, Leviton L. *Foundations of program evaluation*. Sage Publications, 1990.

Robust evaluation tells us not only whether an intervention (the term we will use throughout to refer to a quality improvement project, programme or initiative) worked, but also why and how – allowing us to learn lessons for spreading successful interventions and developing new ones.

Evaluation that is done inadequately, or not done at all, can render an intervention at best a wasted effort, with improvements only realised at local level. At worst, evaluation can lack credibility, especially if there is a bias towards emphasising success and ignoring failure, which can undermine efforts to improve patient care.

Why do an evaluation?

1

Besides demonstrating that an intervention has been a success – or, equally importantly, did not achieve what was planned – many other things can be learned. Asking the apparently simple question, 'Does it work?', can lead to more complicated and useful considerations. For example, if it is working now, will it continue to work in future? Will it work somewhere else? Who did it work for? Did it work in the way we thought it would? What made it work?

By addressing these wider questions, an evaluation can help you make informed decisions about whether your intervention has made beneficial changes in the most effective way. Evaluation can also benefit others – without evaluation, improvements may only be known about and understood in the locality where they took place.

Evaluation captures insights that might otherwise be lost over time and generates new knowledge, so others interested in improving quality of care can benefit from lessons learned. Communicated in the right way, this can help steer the development of new policies and new ways of working.

Evaluation also provides the sort of information that is useful when it comes to determining what will happen to an intervention in the future. For example, if it is successful on a small scale, how could the new ways of working be spread elsewhere?

Evaluation can serve a number of purposes:

- For the intervention team, evaluation is a means of learning whether a desired change has been achieved, whether it represents an improvement, whether it caused problems, whether it has served patients' interests, how the intervention worked and whether it is sustainable. If the evaluation is done during rather than after the intervention, it could provide feedback on a continual basis to allow those involved to reflect on and review the improvement process in real time.

- Depending on the design of the evaluation (see section 3, 'What are the design considerations for an evaluation'), it can be an opportunity for patients' and patient representatives' perspectives to be heard and taken into account. Evaluation findings can be shared with user groups such as patients, carers and the public to show how NHS organisations are working to improve quality of care and change practices for the better.

- For funders or budget holders, it can provide feedback on whether the provision of financial and human resources needed for the intervention was justifiable, thus ensuring

transparency and accountability. It can help them to decide whether or not to fund future work and gain 'buy-in' from others.

- For those undertaking improvement work, evaluation can be a resource for learning and sharing knowledge about what others have found to work, and what they have found to be difficult or unsuccessful. This can include whether and how learning within a particular context might be applied more generally.

With all these different stakeholders involved, you will need to take steps to resolve varying expectations about what an evaluation can and can't do, and negotiate priorities, so they can be clearly communicated to the evaluation team. From the start, stakeholders should also be clear on what they should and are able to provide to help deliver a good evaluation: for example, people's availability to take part or access to data and background documents. There are different ways this communication can happen, such as joint working on bids from the outset, or intervention and evaluation teams being represented at each other's strategic or management meetings and sharing written information about the evaluation with relevant parties.

What are the different types of evaluation?

2

There are many types of evaluation. Here, we give a brief overview of two of the most commonly used types (summative and formative) and two fairly new ones (rapid cycle and developmental). There are no firm rules about which approach to take and your choice will depend on a number of factors:

- what you and the intervention team hope to learn from the evaluation
- what the different stakeholders' needs are
- how long you have to carry out the evaluation, *and*
- what your budget is.

Summative

A summative evaluation can be seen as a 'summing up' of the overall effect of the intervention. It is often carried out at the end, when all the data are available to help the evaluation team to determine whether it has been a success or not, often against stated goals.

This type of evaluation might show whether the intervention worked and met its objectives, what improvements, if any, it created, and how the benefits compared to the costs. It is useful for judging the overall worth and significance of an intervention in a way that helps senior managers, budget holders or funders decide whether it should be continued,

modified or even adopted on a wider organisational level. A summative evaluation works best if the intervention and the environment in which it is carried out are unlikely to change, or at least not substantially, during the evaluation.

Formative

A formative evaluation is designed to help form or shape an intervention. It is used as the intervention evolves and can provide information about how best to revise and modify the work taking place. It can help people to explore not only whether improvement has been achieved, but also how it has occurred in their particular environment.

The data from a formative evaluation is likely to be both quantitative (numerical data through statistics, surveys, questionnaires and structured interviews) and qualitative (semi-structured or unstructured interviews, focus groups, observations and document analysis). The data can be used to develop the intervention, fixing implementation problems so that it is more likely to be successful.

In reality, any evaluation is likely to have both summative and formative elements, to address whether something works and understand why it produces specific results for future iterations.

Rapid cycle

Rapid cycle evaluation is an example of formative evaluation which aims to use 'single loop learning' – where the goals are treated as being relatively fixed, but details about how to obtain these goals might be refined. Methods are used to determine on a regular basis whether an intervention is effective, and enable people to continuously improve their interventions by experimenting with different adaptations.

Rapid cycle evaluation can be used to assess large-scale changes, such as providing patients with a new set of services, as well as small changes such as rewording letters that encourage patients to take a particular course of action. For example, in the US, the Center for Medicare and Medicaid Innovation aimed to provide Pioneer Accountable Care organisations with ongoing feedback to support continuous quality improvement. It did this by updating its estimates about the effectiveness of these interventions on a regular basis (assessed against a matched control group).[*]

---

[*] Shrank W. The Center For Medicare And Medicaid Innovation's blueprint for rapid-cycle evaluation of new care and payment models. *Health Affairs* (Project Hope). 2013; 32(4);807–12.

Developmental

A developmental evaluation involves 'double loop learning' – where the innovation theories and assumptions are revised over time, with the result that the goals of the intervention might also be changed. This type of formative evaluation also facilitates real-time, or close to real-time, feedback to the intervention team. It assists with trying out new ideas, documenting activities and their short-term consequences, identifying processes and outcomes as they emerge and helping people to make sense of them. This allows ongoing development of the intervention, theories of change and occasionally the aims of the initiative. It is best suited to an improvement initiative that is looking at innovative solutions or social change and may be needed in a complex or uncertain environment.

A developmental evaluation requires the intervention team and those designing or doing the evaluation to work closely together with mutual trust, be open to flexible methods of working and be able to synthesise potentially conflicting data. Without this openness and flexibility, the evaluation will have limited value in helping the intervention develop effectively.

What are the design considerations for an evaluation?

3

You need to be clear at the outset of planning an evaluation what is to be evaluated and what you (and your stakeholders) want to learn. These are the core issues that need to be addressed by your research questions, which then drive the design. The study design has to use the research methods and data that are most suitable to provide answers to the research questions. Developing a theory of change or logic model can be helpful in determining your research questions, as well as developing the intervention itself, as seen in box 2.

Box 2: Theories of change and logic models

A 'theory of change' is a narrative approach that, working back from the desired change, methodically sets out what needs to improve, what changes are likely to deliver the improvement and what action is required for the changes to happen.

A logic model is more diagrammatic, working forward from the actions that need to take place to drive the change that is likely to result in the desired improvement. These approaches are both based on structured thinking processes to understand how and why a desired change is expected to come about (as a result of an intervention).

Whichever approach is used, the value is in the process of working with your team to think through the details of how you expect your intervention to work and explore the assumptions that lie beneath this.

For either a theory of change or a logic model, it is helpful to:

- **keep it simple** (but not simplistic), using language that is clear and easy to understand to a wide range of stakeholders and on a single sheet of paper or screen view

- **keep it relevant**, by reminding people of the practical benefits relevant to their day-to-day work

- **keep it up to date**; it will be a 'live' record that may go through several iterations. This will then tell a clear story of how and why your intervention changed. Schedule regular opportunities for updating your theory, for example project review meetings.

For more detailed information about these approaches refer to the resources section of this guide or visit www.theoryofchange.org

Your answers to the following questions can help to shape your evaluation study design (sometimes also referred to as an evaluation framework or evaluation protocol).

What is the research question (or questions) to be answered?

This will be centred on whether the intervention fulfilled its intended objectives, with sub-questions such as:

- How was this achieved?
- What resources did it take?
- What unintended results were there?

It is important that the research question does not necessarily jump directly to a focus on patient outcomes. This may be what the intervention is ultimately working to improve but demonstrating a change in health outcomes for patients may require a large sample, together with a considerable period of time (several years in many cases) in order to show any change with a high level of statistical robustness. Many other measures will also be relevant, such as improved processes of care and efficiencies for use of staff time and other resources.

For example, it was important for the evaluation of the Health Foundation's Engaging with Quality in Primary Care programme to capture the rich information about how primary care staff were working with the challenges and opportunities of implementing improvement, in addition to benefits for patients.[*]

What type of intervention is being evaluated?

As there are countless ways of evaluating a wide range of interventions, the most suitable method of evaluation will depend on the nature of the intervention itself.

For example:

- Is it dealing with a single quality improvement issue in one context, or is it multifaceted, involving several simultaneous changes?

- What stage of development is it at? Has it been previously used and refined, therefore 'stable', or likely to evolve or change during the course of the evaluation?

---

[*] See: www.health.org.uk/publications/involving-primary-careclinicians- in-quality-improvement

- What is currently happening in the context or system in which it is being implemented or used? (Consider issues such as organisational structure, relocation of services, central policy initiatives, changes in staff roles.)

- How many locations are affected by the intervention, and is it possible to include them all in the evaluation or will it be necessary to take a sample?

An evaluability assessment may be useful in determining whether the intervention is ready to be evaluated (see box 3).

Who are your stakeholders and what are their priorities?

It is important to take into account your stakeholders and consider what evaluation designs will be suitable to address their needs. For example, who is affected by the intervention (patients, colleagues in one department or many departments, people in other organisations)? Who involved in the intervention's implementation will be interested in the evaluation (leadership of your organisation, commissioners of the service, funders, data collectors and analysts)? Who else might find the results of the evaluation useful (people running similar services, policymakers, academic researchers)?

Box 3: Evaluability

The idea of 'evaluability' can be helpful when designing the evaluation, to help understand the nature of the intervention and how to approach the evaluation. Some indicators that an intervention is ready to be evaluated include:

- being very clear on what change you are trying to achieve

- assessing that the intervention methods and implementation plan are realistic and acceptable to the people involved.

You should ask yourself about the evaluability before the start of an intervention and again after any major changes. If the answers show that the implementation is still in development, a summative evaluation may not be appropriate. A carefully planned study to monitor the piloting of the intervention may produce more useful, timely information to further develop the intervention.

What are we comparing our intervention with?

4

At its heart, evaluation is a process of comparison. The design of your evaluation will also be informed by what you decide to compare your intervention with: itself over time, to a comparator group, or to what is known as a control group in a randomised trial. Indeed, it might not even be possible to formulate the questions that will be addressed by the evaluation without an idea of the comparison – the intervention led to improved outcomes compared to what?

The kind of comparison you make will be determined by the questions that are being addressed, while practical issues such as budget and expertise, as well as the implementation and delivery context, will also be relevant. Each approach also has different strengths and weaknesses. We give an overview of these, but they will need to be thought through in detail as part of your evaluation design.

Comparing the intervention with itself

Comparing the intervention with itself over time in a before-and-after or time series analysis is relatively straightforward to do as long as the 'before' or baseline data is collected prior to the start of the intervention. However, before-and-after studies can give very misleading results.

There are two common reasons why this can happen:

- Firstly, things may get better at your clinic/practice/hospital following the intervention, but that may happen elsewhere too. Hospitals have found, for example, that mortality rates or infection rates fall after interventions, but when they then compare themselves to similar hospitals they find they too have improved without that intervention.

- The second reason why we may be misled is what is called 'regression towards the mean'. To understand this, imagine someone with a chronic condition in which their health goes up and down over time. If they are particularly ill one day they will probably be back nearer to their average condition the next day. However, if they were taken into hospital on their 'bad' day and given a new treatment, the fact that they were better afterwards might not have been because of the treatment but because they were just going back to their average health (regressing from where they were towards their mean).

Other limitations of before-and-after studies are that they can only be said to relate to the context and group treated, so it is harder to generalise from them.

Comparison with a comparator group

Comparing the effects of your intervention with a comparator group which has not benefited from the intervention will help you to understand whether any change can be attributed to the intervention or other developments that were taking place at the time, such as changes in staff or policy. This will increase the level of internal validity.

It is important to recognise the risk of selection bias. This happens when there is a difference between the intervention group and the group chosen as the comparator. Bias can creep in for many reasons. For example, when working with patients on a new way of shared decision making, health professionals may select patients who are easier to get on with for the intervention. As it takes time and effort to introduce something new, there may be a temptation to run an intervention on wards or clinics which are less busy than others; this decision itself leads to bias. There is also bias that can happen in ethnically diverse communities because some groups may require specialist skills to access, perhaps because of language difficulties.

The most common way to deal with selection bias is by randomly allocating patients or groups or clinics (we deal with this in the next section). There are other ways of dealing with selection bias – for example, it may be possible to fit regression models or select a matched subset of comparison patients who are similar to the intervention group. However, these methods must be applied and interpreted carefully.

Randomised control trials

Randomised control trials (RCTs) will give you the most confidence that any change can be attributed to the intervention. They can be costly and complicated to design and are dependent on specific factors being in place, such as being able to deliberately assign whether people are subject to an intervention or not. Furthermore, an RCT can sometimes affect both the implementation of the intervention and the context in which it operates, reducing the generalisability of the findings in comparison with other evaluation methods. In some circumstances randomisation may not be possible for ethical or practical reasons. Hospitals, for example, choose electronic health record systems according to their own needs and budget and this is too important for them to accept being randomised to one system or another.

How does evaluation differ from other forms of measurement?

5

Terms such as 'evaluation', 'audit' and 'performance measurement' are sometimes used interchangeably, but there are differences, if subtle.

Audit

An audit is an investigation into whether a service or activity is in line with agreed standards, to see if and where improvements can be made. The standards used can be determined by the service providers themselves or externally. An audit could be conducted at a national, regional or local level, and the knowledge produced by an audit usually applies only to the particular context in which it was conducted – there is no attempt to generalise.

Example:

In recent years national clinical audits have been carried out across a wide range of medical, surgical and mental health conditions. The National Diabetes Audit measures the effectiveness of diabetes health care against NICE Clinical Guidelines and NICE Quality Standards in England and Wales to drive changes and improvements in the quality of services and health outcomes for people with diabetes.[*]

---

[*] See www.hscic.gov.uk/nda

Performance measurement

This uses data to try to determine the progress of a particular intervention or service against a set of targets or objectives. Unlike an audit, data are usually collected and used at regular intervals to report progress to management teams; for example, as part of governance processes or for staff appraisals. Like an audit, data from performance measurement might be used to identify areas of concern in a service or where there is potential for improvement.

Example:

The NHS's A&E Attendances and Emergency Admissions collection looks at the total number of attendances in any week for all A&E types and the percentage of patients who were discharged, admitted or transferred within four hours of arrival. Data are then used by provider organisations such as NHS trusts to see how they compare to the national target of 95%.

How does evaluation differ?

By contrast, evaluation is less about strict protocols, predetermined standards or routine day-to-day management and more about a practical assessment of the implementation and impact of an intervention. It is conducted in a spirit of discovery rather than management or monitoring. It is concerned with developing understanding and supporting more strategic judgement and decision making, such as whether and how an intervention should continue, and continue to be funded.

Evaluation can draw on the routinely collected data that is used for audit, performance management and other purposes. It is desirable to use data that is already available to reduce the burden of additional data collections from people involved in your intervention. For example, you may already be using measures that test whether change is happening in processes, if outcomes are changing or if key performance indicators are shifting, as you monitor the effects of your intervention.

The data from whatever measures you are using will also be of great value in the evaluation to contribute to a deeper understanding of whether and how the intervention worked. For example, Sheffield Teaching

Hospitals NHS Foundation Trust monitored geriatric medicine bed occupancy rates to understand how interventions to improve patient flow were working.[*]

An evaluation will need to consider whether the routine data available can be used as a robust measure of the impact of the intervention. Data collected for one purpose, say day-to-day management of a hospital ward, may not be suitable to distinguish whether a specific component of an improvement intervention is having the planned effect. Biases can result if the analysis does not carefully reflect the mechanism by which the data were generated.

---

[*]   For details of this work in Sheffield, see the learning report
        Improving patient flow: www.health.org.uk/publications/
        improving-patient-flow

What practical issues should we consider?

6

In considering some of the bigger issues outlined elsewhere in this guide, it may be that you neglect some of the smaller, practical issues. These can be important in preventing delays to the process as well as keeping within budget.

Milestones

What decision-making points (such as spreading the intervention to new teams or contract renewal meetings) exist in relation to the intervention? What needs to be known before such decisions can be made? When do they need to be made? What length of time is required to explore the evaluation questions you are interested in? When might you realistically expect to see the changes you are looking for? The best way to align evaluation activity to support important milestones is to consider the evaluation plan as you start planning your intervention.

Obtaining ethical approval

Generally, if you are interviewing or observing patients or the public or looking at confidential documents such as patient records as part of the evaluation, you will need to obtain ethical approval. Your organisation or funder may have a research or governance manager who can guide you through the process. It can take several months so you should start as early as possible.

You will need to prepare, among other things, a protocol (a summary of the evaluation and how it will be carried out), a plan for the evaluation, information about the evaluation for all participants, consent forms and an explanation of the procedure for obtaining consent, and details of the skills and qualifications of the evaluation team. All information for participants should be written in plain English, and other languages as appropriate.

Evaluations of service improvements that use routine and anonymised data do not usually require ethics committee approval. You should seek guidance from your local research ethics committee. The NHS Health Research Authority also provides a useful guidance leaflet, which recognises that decisions about the need for ethics committee approval are not always clear.[*]

Collecting and analysing data

Depending on what you hope to learn from the evaluation you may have to gather a large amount of quite complex data. The time required for collection and analysis is almost always underestimated. The data is likely to be a mix of quantitative and qualitative methods. Access to both types of data requires formal (eg approval from trusts, security

[*] www.hra.nhs.uk/documents/2013/09/defining-research.pdf

clearance) and informal (eg diary time to conduct interviews) negotiation at different levels within participating organisations.

Data quality is often poor and experience has shown that many NHS data sets are incomplete, not well coded, are not up to date or can be difficult to access. It will take time to access, clean and quality-assure data, and test levels of completeness and the implications of this for analysis.

Doing the paperwork

Depending on the way in which the evaluation is being undertaken and who is carrying out or funding the work, there will be different requirements for formal agreements between parties. If you are commissioning an external evaluation team you will need to draw up a formal written agreement that sets out clear deliverables and milestones, particularly where decisions need to be made, and a procedure for revising the evaluation design if an intervention changes.

If there are many stakeholders, or different stakeholders collecting data for different purposes (for evaluation, for learning and improvement or for monitoring), it may be worth drawing up a

concordat or Terms of Reference to establish clear roles and responsibilities to help manage any potential tension, duplication or uncertainty.[*]

There will also be data protection guidelines to review and adhere to, informed consent paperwork to draw up and complete and possibly site-specific paperwork to secure access to buildings and data.

Building a culture of learning

Remember that all forms of evaluation have the potential to be seen as threatening to the intervention team, their colleagues and other stakeholders. It is beneficial if evaluation takes place in a learning climate where staff are open to constructive feedback and change and are not fearful of the consequences of negative results.

---

[*]   For example, as part of phase two of the evaluation of the Health Foundation's Safer Clinical Systems programme, the University of Leicester's SAPPHIRE research group developed a concordat that clarified and documented expectations, including the respective roles of evaluator and intervention teams and data collection resourcing and responsibilities.

See: Brewster L, Aveling EL, Martin GP, Tarrant C, Dixon-Woods M, and The Safer Clinical Systems Phase 2 Core Group Collaboration & Writing Committee. What to expect when you're evaluating healthcare improvement: a concordat approach to managing collaboration and uncomfortable realities. *BMJ Quality and Safety* (in press 2015) http://qualitysafety.bmj.com

When should we start and finish an evaluation?

7

A well-designed intervention will include provision for evaluation from the outset. Building in data collection, feedback and measurement of outcomes and impact from the start can ensure that the evaluation is integrated into the design and execution of the intervention. This positions the evaluation as an integral part of the process and delivery rather than an interference competing for tight resources.

You will need to be clear about what data you need: this is one of the most underestimated aspects of an evaluation. If possible, data that are routinely available should be used, but sometimes specific data is needed to address the questions an evaluation is posing. The sooner evaluators are involved, the greater the influence they are likely to have over data collection and access, potentially increasing the options for the evaluation design.

If you need to collect additional information there may be a limited opportunity to do so; some baseline information may only be available before an intervention starts, and cannot be recaptured once an intervention is implemented. For example, if you wanted to know how a new person-centred care intervention influenced patients' experience over time, it would be necessary to collect data on their experience before the intervention was implemented to compare

with their later experiences. Data may need to be collected even sooner than expected if, for instance, a rumour about a proposed change starts to influence behaviour before the intervention formally begins.

The end point of an evaluation should largely be determined by the length of time it is likely to take to see the changes that the evaluation is trying to capture and measure. Will the evaluators be collecting data for long enough for you to see the changes you are hoping to measure with a sufficient degree of reliability? It is possible that the initial changes you expect to see within the first few months of implementation, and which may be possible to capture through evaluation, may differ from your longer-term ambitions for your intervention.

The length of time needed to look for expected change has to be balanced against short-term needs: some stakeholders may need to have information within a particular timeframe to, say, inform policy or justify funding decisions. It is important to manage expectations from the start, identifying intermediate outcomes that can be reported early on while being very clear about the limitations of what indicative findings can and cannot say about eventual outcomes, which may only become obvious a year or more after the intervention has been implemented.

How do we cope with changes in the intervention when the evaluation is underway?

8

Many interventions are a step into the unknown and do not always go according to plan. This is not necessarily unexpected or unwelcome. It is not unusual for interventions designed for quality improvement purposes, particularly those that are still evolving, are complex, or are occurring in dynamic environments, to change during the course of an evaluation.

These changes may be minor (small delays in plans, changes in staff involved in the intervention) or they may be fundamental to the overall evaluation design (a change in what the intervention is trying to achieve or the context in which the intervention is working). To help manage change it can be useful to:

- Understand the likelihood of change when designing an evaluation. This is a helpful first step as it will influence the type of evaluation chosen. For example, a developmental approach may be more appropriate when it is understood that the intervention will be continuing to evolve and develop. Balancing flexibility and robustness in evaluation design means developing a plan which is realistic about what is known, what you know might change and what some of the options might be for addressing potential changes.

- Ensure good channels of communication between intervention and evaluation teams. While it is not possible to avoid difficult conversations, good channels of communication mean the implications of change can be discussed at the earliest opportunity, and expectations about what the evaluation will then be able to deliver can be managed. Examples include: evaluation and intervention teams attending and having a degree of involvement in each other's meetings; sharing documents and project management plans through online tools such as Dropbox, Huddle and Trello; and facilitating access for evaluators to observe the intervention in practice.
- Align the evaluation design to the intervention design, so the evaluation model can be regularly reviewed in light of changes to the intervention. This will help you to be more flexible and plan for change more easily.

Changes in your intervention can lead to positive opportunities for an evaluation, for example to refine data collection to be more focused and less burdensome and to renegotiate its objectives and the use of resources. If an intervention

undegoes fundamental changes in aim, content or implementation, there may not be the means to capture these in the current evaluation study.

A judgement will need to be made about whether to stop, continue or adjust the study. There should be particular consideration of the level of detail needed and achievable.

It is worth remembering that unanticipated changes may contain important lessons and should be recorded and reported in the spirit of transparency and learning.

Should we do the evaluation ourselves or commission an external team?

9

There are advantages and disadvantages to carrying out an evaluation yourselves or commissioning an outside team to do it for you.

Internal evaluation

An obvious advantage is that this choice is likely to be cheaper and may suit – indeed, may be the only option for – those on a limited budget. It may allow greater control over the process in terms of collecting data to monitor the progress of the intervention and making necessary changes to the intervention design immediately. It may also be more inclusive, encouraging engagement and participation from a wider range of people within the organisation, and helping them to develop new skills, knowledge and interests.

However, this approach may require more skills, experience and resources than your organisation has at its disposal in terms of project management, research expertise, IT skills, data analysis and even communication. There may be a conflict in prioritising evaluation over other work commitments, with day-to-day tasks needing to take precedence. In addition, it may be harder for an internal team to develop and retain a degree of independence from the intervention as they may

have relationships with or pre-existing opinions of the stakeholders, structures or processes involved in the study. There is a danger that objectivity may be compromised (or be perceived to be compromised): an internal team may have a vested interest in the success of an intervention or have a preconceived idea of how it is supposed to work, and this may affect the reporting of results, even unconsciously.

Externally commissioned evaluation

A specially commissioned team can be expected to have the skills and expertise in evaluation techniques. This may enhance the reputation of the study with external stakeholders such as funders or peer reviewers, who may feel the findings have more credibility. An external team may have the experience to work more efficiently and effectively through experience of applying different types of design and method. Also, by drawing on and contributing to relevant theory and knowledge of the subject matter, they can better achieve synergy between the existing evidence base and the evaluation. By taking a more independent stance, their findings may be more nuanced and perceptive than an internal team's.

On the other hand, an external evaluation could prove more expensive and time-consuming as the team will need to understand the context and the aims and objectives of the intervention. The need to communicate this to the evaluation team, and develop their understanding of tacit knowledge – that which is not formally codified and can be more readily observed from practice – will take time, as will setting up the processes for supporting access to buildings and data, for instance. Moreover, an external team may arouse some suspicion and resentment among participants who may feel spied on, particularly if external funding is involved, and the relationship between the participants and evaluators will need to be managed carefully.

As an evaluator from one of the Health Foundation programmes has noted:

'Our reflections about the process so far have also centred on the need to achieve a common "language" across the implementation and evaluation teams – this has required much checking and rechecking about our shared understanding of project aims, goals and timescales.'

Joint working

There is also the option of a joint evaluation, or division of roles, with, say, an internal team managing or at least supporting the process, perhaps undertaking data capture, and an external team assuming responsibility for a literature search or data analysis and offering advice on the preparation and presentation of the findings.

How do we communicate evaluation findings?

10

Communicating the findings of an evaluation should be driven by a clear understanding of who or what is to be influenced in terms of stakeholders involved and decisions to be made. You need to build in time and resources to make the findings relevant and accessible to your target audiences, who are the people you want to influence.

A plan to develop different methods of communication can then be devised to suit the needs and expectations of different stakeholders. This can be started at the beginning of an evaluation, and updated as a live document, as who will be interested and what they will be interested in becomes clearer. Planning should also clarify intellectual property rights and how to manage the sequence of publications. For example, if an evaluation team is seeking publication in peer-reviewed journals, this may have implications for the publication of results through other channels such as an organisation's website.

All those involved in the intervention should get feedback based on the findings. To ensure they remain engaged, internal stakeholders should not be neglected in favour of external stakeholders as the results of the evaluation may feed into future organisational decision making.

For people concerned with the practical running of an intervention, feedback may be needed regularly to adapt and improve the intervention, making interim reports and less formal feedback mechanisms vital in informing key decision points with a summary of progress so far, as well as initial findings. This is particularly important if the evaluators gain insights that the findings are particularly sensitive or contradictory to general perceptions of the intervention's effects.

There are formal guidelines and conventions for reporting. For example, over the past two years, with support from the Robert Wood Johnson Foundation in the US and the Health Foundation in the UK, the Standards for Quality Improvement Reporting Excellence (SQUIRE) guidelines have been revised and updated.[*]

While this level of formality and depth of reporting should be proportionate to the size or purposes of an evaluation, formal guidelines can provide useful insight into the kind of information that should be easy to access in some form.

---

[*]   The final draft is available for review and public comment at http://squire-statement.org

There are many other perhaps more creative and helpful ways to communicate findings alongside a formal report. These could include high-level briefings or slide decks of key findings (for senior managers/funders), academic papers and conference presentations or posters (for other improvement specialists or researchers who could learn from your work), press releases or news stories (for PR or marketing purposes), YouTube videos or e-newsletters (for all stakeholders) and intranet and staff meetings (for colleagues).[*]

It is important to be transparent about all of the findings, including those that are less positive, and the data and processes that led to them, even if these are reported in a separate 'technical appendix', to keep the main findings to an accessible summary.

Box 4 contains the information that should be documented and made available, usually in a report.

---

[*] See also the Health Foundation practical guide *Using communications approaches to spread improvement* www. health.org.uk/commsforspread

Box 4: Example contents of a traditional evaluation report

- **Executive summary:** a very brief overview of the intervention that was evaluated; the main findings and/or recommendations.

- **Introduction:** a description of the intervention being evaluated, what it was intended to achieve and how, along with brief information on the context where it took place; a summary of the evaluation plan, who did it, time period, funders, main research questions.

- **Background:** detail on what is being evaluated; previous evidence/ research on the area; relevant policy and practice context.

- **Method:** research questions for the evaluation; theory underpinning approach; sampling; data collection and analysis; limitations/caveats; changes in method and reasons why; theory of change visuals (these can work well in the main body of the text if they are clear); description of formula used in analysis; detailed analysis charts; example of research tools – survey, interview schedules; full description of sample.

- **Findings and discussion:** reports can present a more traditional pure description of findings in one section, followed by a discussion. Reports can flow more easily for the reader if these two elements are combined, perhaps structured around the main research questions. It is important to present all of the findings in the main body of the report, not only the positive findings.

- **Recommendations and lessons for future:** a separate section that discusses the implications of the findings for the future (sustainability and development) and may make specific recommendations.

- **Appendices:** might include a glossary if technical terms are used; list of participants; additional methods detail (as mentioned above); references; detailed budget information if relevant and so on (length will vary).

- **Thanks:** to people who have contributed to the process.

Resources

The following list contains a collection of useful websites, articles, webinars and other guidance on various aspects of evaluation. While this is by no means an exhaustive list of resources, we hope that it is a starting point for locating further information to help you plan and undertake successful evaluations.

General evaluation resources

Resource	Available from
Article: *Bridging the ivory towers and the swampy lowlands; increasing the impact of health services research on quality improvement* Martin Marshall	**International Journal for Quality in Health Care.** http://intqhc.oxfordjournals.org/content/early/2013/10/17/intqhc.mzt076
Book: *Evaluating improvement and implementation for health* John Øvretveit	**Open University Press.** www.mheducation.co.uk/9780335242771-emea-evaluating-improvement-and-implementation-for-health
Document: *Developing and evaluating complex interventions: new guidance* Peter Craig, Paul Dieppe, Sally Macintyre, Susan Michie, Irwin Nazareth, Mark Petticrew	**Medical Research Council.** www.mrc.ac.uk/complexinterventionsguidance
Document: *Evaluation Flash Cards: embedding evaluative thinking in organizational culture* Michael Patton	**Otto Bremer Foundation.** www.ottobremer.org/sites/default/files/fact-sheets/OBF_flashcards_201402.pdf
Document: *Impact evaluation glossary*	**International Initiative for Impact Evaluation (3ie).** www.3ieimpact.org/media/filer/2012/07/11/impact_evaluation_glossary_-_july_2012_3.pdf

Resource	Available from
Document: *Process evaluation of complex interventions* Graham Moore, Suzanne Audrey, Mary Barker, Lyndal Bond, Chris Bonell, Wendy Hardeman, Laurence Moore, Alicia O'Cathain, Tannaze Tinati, Danny Wight, Janis Baird	**Medical Research Council.** www.populationhealthsciences.org/ MRC-PHSRN-Process-evaluation-guidance-final-2-.pdf
Videos: *Professor John Øvretveit shares evaluation expertise in exclusive Q&A videos* John Øvretveit	**The Collaboration for Leadership in Applied Health Research and Care (CLAHRC) Greater Manchester.** http://clahrc-gm.nihr.ac.uk/2014/07/ prof-john-ovretveit-shares-evaluation-expertise-in-exclusive-qa-videos/
Webinar: *The benefits and challenges of evaluating improvements* Nick Barber and Laura Leviton	**The Health Foundation.** www.health.org.uk/multimedia/video/ benefits-and-challenges-of-evaluating-improvements
Website: *Better Evaluation: An international collaboration to improve evaluation practice and theory*	**Better Evaluation.** http://betterevaluation.org/start_here
Web resource: About monitoring and evaluation	**Charities Evaluation Services.** www.ces-vol.org.uk/about-performance-improvement/about-monitoring-evaluation

What are the different types of evaluation?

Resource	Available from
Article: *How to study improvement interventions: a brief overview of possible study types* Margareth Crisóstomo Portela, Peter J Pronovost, Thomas Woodcock, Pam Carter, Mary Dixon-Woods	**BMJ Quality and Safety.** http://qualitysafety.bmj.com/content/early/2015/03/24/bmjqs-2014-003620.full
Document: *A developmental evaluation primer* Jamie Gamble	**J.W. McConnell Family Foundation.** www.mcconnellfoundation.ca/en/resources/publication/a-developmental-evaluation-primer
Document: *A practitioner's guide to developmental evaluation* Elizabeth Dozois, Marc Langlois, Natasha Blanchet-Cohen	**J.W. McConnell Family Foundation.** www.mcconnellfoundation.ca/de/resources/publication/de-201-a-practitioner-s-guide-to-developmental-evaluation
Document: *Handbook on impact evaluation: quantitative methods and practices* Shahidur Khandker, Gayatri B. Koolwal and Hussain Samad	**World Bank.** http://elibrary.worldbank.org/doi/book/10.1596/978-0-8213-8028-4ID=000333037_20091210014322
Website: *Cognitive Edge:* A network of practitioners and source of resources for working with complexity and narrative	**Cognitive Edge.** http://cognitive-edge.com/
Web resource: *Online guide: Everything a qualitative health researcher needs to know* M Nakashian	**Robert Woods Johnson Foundation.** www.rwjf.org/en/research-publications/find-rwjf-research/2008/05/online-guide.html

How does evaluation differ from other forms of measurement?

Resource	Available from
Web resource: *Deciding if a study is research, audit, development or service evaluation*	**NHS Research and Development Forum.** www.rdforum.nhs.uk/content/ resources/#Deciding

What are the design considerations for an evaluation?

Resource	Available from
Article: *Demystifying theory and its use in improvement* Frank Davidoff, Mary Dixon-Woods, Laura Leviton, Susan Michie.	**BMJ Quality and Safety.** http://qualitysafety.bmj.com/content/ early/2015/01/23/bmjqs-2014-003627. full
Article: *Recommendations for Evaluation of Health Care Improvement Initiatives* Gareth Parry, Andrew Carson-Stevens, Donna Luff, Marianne McPherson, Donald Goldmann	**Academic Pediatrics.** www.academicpedsjnl.net/article/ S1876-2859%2813%2900099-5/ fulltext
Document: *A practical guide for engaging stakeholders in developing evaluation questions* H Preskill and N Jones.	**Robert Woods Johnson Foundation.** www.rwjf.org/en/library/research/ 2009/12/a-practical-guide-for-engaging- stakeholders-in-developing-evalua.html
Web resource: *Developing a robust protocol design*	**NHS Research and Development Forum.** www.rdforum.nhs.uk/content/ resources/#Robust
Webinar: *Achieving synergy between designing and reporting quality improvement projects* Kaveh Shojania	**The Health Foundation.** www.health.org.uk/multimedia/video/ achieving-synergy-between-designing- and-reporting-quality-improvement- projects
Website: *Center for Theory of Change: a non-profit organisation established to promote quality standards and best practice for the development and implementation of Theory of Change.*	**Center for Theory of Change.** www.theoryofchange.org

How do we communicate evaluation findings?

Resource	Available from
Article: *Evidence based guidelines or collectively constructed 'mindlines'? Ethnographic study of knowledge management in primary care* John Gabbay, Andrée le May	**BMJ.** www.bmj.com/content/329/7473/1013
Blog: *Infographics to make your evaluation results go viral* Joitske Hulsebosch	**Better Evaluation.** http://betterevaluation.org/blog/infographics_to_make_your_eval_results_go_viral
Webinar: *How to write about quality and get published* Mary Dixon-Woods	**The Health Foundation.** www.health.org.uk/multimedia/video/how-to-write-about-quality-and-get-published-improvement-science-webinar
Website: *Standards for Quality Improvement Reporting Excellence (SQUIRE) Guidelines*	**SQUIRE.** www.squire-statement.org

Examples of evaluation communication materials

Resource	Available from
Information leaflet: *Evaluation of the feasibility of using the Patient Activation Measure in the NHS in England* The University of Leicester	**The Health Foundation** www.health.org.uk/evalexample-infoleaflet
Newsletter: *VERDIS Newsletter 2014* The University of Nottingham	**The Health Foundation** www.health.org.uk/evalexample-newsletter

The Health Foundation is an independent charity committed to bringing about better health and health care for people in the UK.

Our aim is a healthier population, supported by high quality health care that can be equitably accessed. From giving grants to those working at the front line to carrying out research and policy analysis, we shine a light on how to make successful change happen. We use what we know works on the ground to inform effective policymaking and vice versa.

We believe good health and health care are key to a flourishing society. Through sharing what we learn, collaborating with others and building people's skills and knowledge, we aim to make a difference and contribute to a healthier population.