The Strategy Unit.

Designing programmes to support impact evaluation

Short report

January 2025

Prepared by:
Mike Woodall
Head of Impact Evaluation
m.woodall@nhs.net

NHS
Midlands and Lancashire
Commissioning Support Unit

Making the best use of impact evaluation

Evaluating programmes in principle is easy. Evaluating them in the real world can be fiendishly tricky. The real world provides a complex environment that destroys many plans at first contact. There are usually multiple reasons why the outcomes you want to achieve might change. Many of these are unrelated to the programme being implemented.

In this short report - part of a series where we draw out lessons from our evaluation work - we look specifically at the challenges of impact evaluation. What is needed for an evaluator to say that a programme caused a change? And how can programme designers take this into account when formulating their plans? Our more detailed guide to evaluation can be found here.

The purpose of impact evaluation is to estimate whether any change in outcomes was 'caused' by the programme in question. This requires the evaluator to control for other reasons that outcomes may have changed - such as changes in population - to isolate the programme impact. Statistical methods are used to calculate whether there is an effect and, if so, whether this effect is significant in statistical terms.

It is important to say at the outset that not every programme needs an impact evaluation. If, for example, the programme aims to spread an intervention that is already known to be effective, then what follows may not apply. But if this not known, then there is a strong case (ethical, financial, often clinical) for designing a programme with impact evaluation in mind.

The final evaluation method is generally decided based on pragmatic issues, such as programme delivery and data availability, rather than technical issues, such as the most robust method. The method should still be robust but that is not the main driver. This means designing the programme to enable evaluation is important.

The design of the programme influences the impact evaluation method

Demonstrating causality has a high benchmark. The programme design can influence the analytical methods available for its evaluation. For example:

- Short programmes leave little time to show impact; and
- Programmes rolled out nationally have no areas without implementation that could be used as controls.

Proper impact evaluation is almost impossible in both cases. Therefore, designing the programme with evaluation in mind will improve your chance of showing the programme has 'caused' the desired improvements.

Building evaluation into the programme design has many benefits. It helps you understand your theory of change, the outcomes you want to improve and the mechanisms that will deliver these improvements. It also supports more robust methods by ensuring they are feasible.

There is no substitute for accessing tailored evaluation advice and support. But it is possible to point to a few main considerations. These include:

- **Geography** – where will the programme be delivered? Should it be rolled out nationally or as local pilot sites?
- **Timing** – what are the delivery timescales? Will all sites start at the same time?
- **Site selection** – where will the intervention be piloted? How do you select the sites to ensure the results can be generalisable? Can controls be identified?
- **Target population** – can the target population be identified from routine datasets? Are the eligibility criteria specific to capture only people eligible for the intervention? Is the sample size big enough to show the expected effect size?

Impact evaluations require data. The need for data runs through all of these considerations. It is required before and after the implementation of the programme, in places with and without the programme. Therefore, data access and data quality are key to all these considerations. New data can only be collected from sites involved in the programme; if this is required, collection needs to be built into the programme design.

Geography

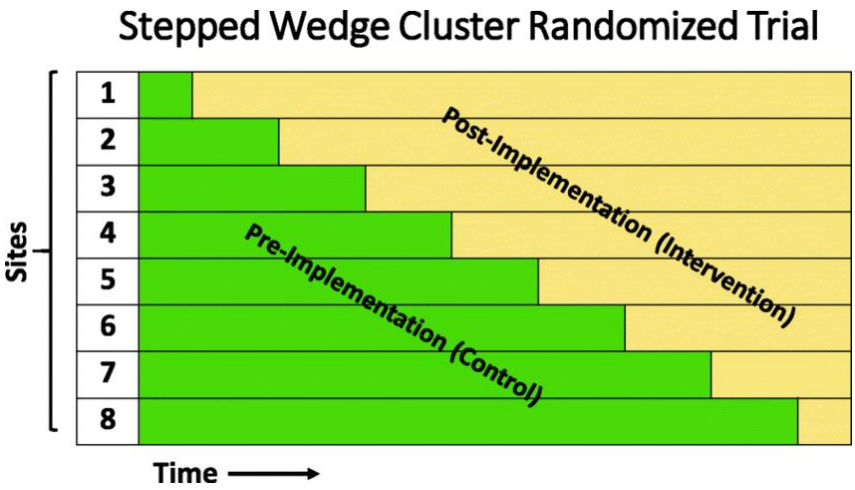
The geography of a programme and the unit of analysis - NHS Trust, Integrated Care Board, Local Authority, etc. - are important considerations in any impact evaluation.

National implementations can be difficult to evaluate. The methods available are limited and may not be able to control for all external factors. Local areas often start implementation at different times, but these dates are not always shared with the programme.

Piloting interventions in different places can increase the methods available and improve the quality of the impact evaluation. Methods that control for confounding factors can then be used. The control sites can be matched using variables, such as deprivation and level of need, to control for the most likely confounders. This increases confidence that any changes were caused by the programme.

This does not mean 'denying' areas a potentially useful service. For example, 'step wedge designs', where pilots that start later can be used as the controls for the earlier pilots, can be used. These are also particularly useful where new data is required as all the pilot sites are signed up to the evaluation. Therefore, it will be easier for the programme to capture data from both the treatment and control sites.

Figure 1: Example of a step wedge design[1]



Timing

Implementation takes time. Single year projects rarely have enough of it to show impact. Evaluations often need to report in year (see our partner report to this one on the challenges with

[1] Source: Dylla, Layne & Douin, David & Anderson, Erin & Rice, John & Jackson, Conner & Bebarta, Vikhyat & Lindsell, Christopher & Cheng, Alex & Schauer, Steven & Ginde, Adit. (2021). A multicenter cluster randomized, stepped wedge implementation trial for targeted normoxia in critically ill trauma patients: study protocol and statistical analysis plan for the Strategy to Avoid Excessive Oxygen (SAVE-O2) trial. Trials. 22. 10.1186/s13063-021-05688-6.

this). This means that the report needs to be complete before the programme has finished. This reduces the data available for analysis.

The main considerations when agreeing project timescales are:

- **Programmes never develop as quickly as planned**. Optimism bias is a well-known problem that is still not well accounted for. Setting up a project and recruiting the right team takes time; most projects don't deliver activity on their first day. This delay in implementation needs to be considered in the programme design
- **Outcomes take time to improve.** Most people who would benefit from the programme will not have an adverse outcome tomorrow. Realistically, when should programmes expect effects to emerge? Tracing through what has to happen for outcomes to change is often illuminating
- **Data isn't available in real-time.** Most datasets have a lag of at least two months. Analysis and reporting then take another two months. This means a 12-month project will only analyse the impact of the first eight months. This is rarely enough time to demonstrate impact.

Several of our recent evaluation reports have included the recommendation to run the impact evaluation again in the future. The interventions were not mature enough to show any impact yet. We have provided the analytical code so it can be run again but it would have been good to have had more time to show impact in the evaluation.

Site selection

If a programme is being piloted, selecting the right pilot sites is important. If all the pilot areas are similar, it will be difficult to generalise the results. For example, if all the pilot sites are in cities, then it will be difficult to understand whether the programme would have an impact in more rural areas.

If sites are selected based on a specific characteristic it can be difficult to identify control sites and generalise the results. For example, if the programme is implemented in all of the most deprived areas, it will not be possible to identify controls with similar levels of deprivation. It will also be difficult to demonstrate whether the same impact would be shown in areas with lower levels of deprivation.

Selecting sites at random or purposively sampling sites can mitigate these issues.

- **Random sampling**. Identifying sites at random may provide a range of sites that cover different populations. Checks can be done following selection to test how representative the sample of the areas you want to include in the programme. Allowing sites to self-select, for example through a tendering process, may also be random but should be checked as other factors, such as better bid writers, might bias selection

- **Purposive sampling**. Identifying different pilot sites that cover different populations will support a robust evaluation. For example, select sites in rural and urban areas or sites in areas with high deprivation and sites in areas with low deprivation.

It is not always possible to randomise your population. A programme we evaluated with a partner was rolled to the areas most at need. Randomised Control Trials (RCTs) had demonstrated the intervention was effective, but the client wanted to understand how the intervention worked at scale. This required a large sample size in a small number of case study sites, which required selecting the sites with the highest level of need. The evaluation showed the intervention had an impact and it is now being rolled out nationally.

Target population

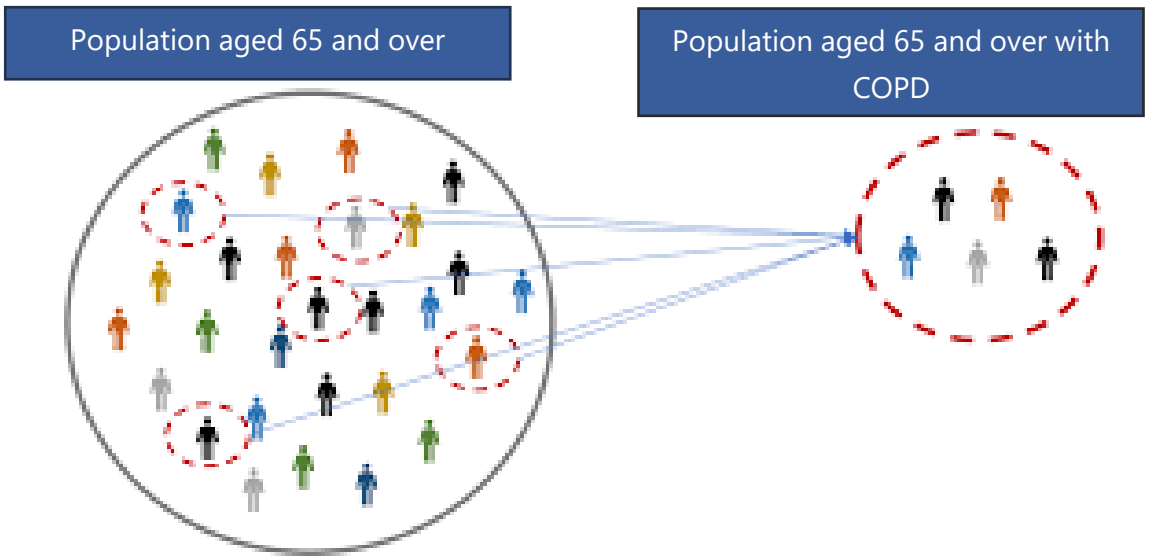
Creating new data flows that track all the participants of a new service is time and resource intensive. They also require extensive information governance processes to allow the data to be used in an evaluation. This means most evaluations in healthcare use routine datasets, such as Hospital Episode Statistics (HES).

It is not possible to identify everyone who engaged in the programme using routine datasets, but it is possible to track everyone who is eligible. Understanding who is eligible for a new service or innovation is important. If you don't know who will benefit from the intervention it is not possible to track their outcomes.

The target population should be accurately defined. This includes any demographic criteria, such as age range, clinical criteria, such as diagnosis or long-term condition, and any geographical or organisational criteria, such as GP network.

Two recent large-scale programmes we evaluated failed to show impact as it was impossible to isolate the target population. A lot of resource was used to design robust methods but ultimately these were flawed because they analysed too many people who could never have benefited from the intervention. This diluted any potential impact and therefore the findings were inconclusive.

Figure 2: Targeting the right population



Ensuring you have the right sample size is essential. Smaller effects are more likely to be statistically significant in larger sample sizes. If the outcome is rare or the sample size is small then you may want to aggregate the data before analysis. This can include aggregating the unit of analysis, for example analysing multiple pilot projects together, or aggregating time periods, for example analysing quarters rather than months. Decisions on how to ensure an adequate sample size should be carefully considered at the start of the impact evaluation.

Why is this important?

Programme design influences the methods available for any impact evaluation. Decisions made at the design stage influence the feasibility of showing the programme's impact. Therefore, better relationships between programme designers and evaluators will improve future NHS evaluations.

The evaluation should not drive the design of the programme, but it is worth considering small changes to the programme design that improve the chance of showing impact. Involving evaluators in the programme design will provide a chance for evaluators to show their value early in the process, rather than after the event.

By working with evaluators to design evaluable programmes the NHS can learn over time rather than repeatedly missing chances to do impact evaluation.

The Strategy Unit
Email: strategy.unit@nhs.net
Web: www.strategyunitwm.nhs.uk