BMJ 2019;365:l2239 doi: 10.1136/bmj.l2239 (Published 20 June 2019)

PRACTICE

ESSENTIALS

Evaluating the impact of healthcare interventions using routine data

OPEN ACCESS

Geraldine M Clarke *senior data analyst*[1], Stefano Conti *senior statistician*[2], Arne T Wolters *senior analytics manager*[1], Adam Steventon *director of data analytics*[1]

[1]The Health Foundation, London, UK; [2]NHS England and NHS Improvement, London, UK

What you need to know

- Assessing the impact of healthcare interventions is critical to inform future decisions
- Compare observed outcomes with what you would have expected if the intervention had not been implemented
- A wide range of routinely collected data is available for the evaluation of healthcare interventions

Interventions to transform the delivery of health and social care are being implemented widely, such as those linked to Accountable Care Organizations in the United States,[1] or to integrated care systems in the UK.[2] Assessing the impact of these health interventions enables healthcare teams to learn and to improve services, and can inform future policy.[3] However, some healthcare interventions are implemented without high quality evaluation, in ways that require onerous data collection, or may not be evaluated at all.[4]

A range of routinely collected administrative and clinically generated healthcare data could be used to evaluate the impact of interventions to improve care. However, there is a lack of guidance as to where relevant routine data can be found or accessed and how they can be linked to other data. A diverse array of methodological literature can also make it hard to understand which methods to apply to analyse the data. This article provides an introduction to help clinicians, commissioners, and other healthcare professionals wishing to commission, interpret, or perform an impact evaluation of a health intervention. We highlight what to consider and discuss key concepts relating to design, analysis, implementation, and interpretation.

What are interventions, impacts, and impact evaluations?

A health intervention is a combination of activities or strategies designed to assess, improve, maintain, promote, or modify health among individuals or an entire population. Interventions can include educational or care programmes, policy changes, environmental improvements, or health promotion campaigns. Interventions that include multiple independent or interacting components are referred to as complex.[5] The impact of any intervention is likely to be shaped as much by the context (eg, communities, work places, homes, schools, or hospitals) in which it is delivered, as the details of the intervention itself.[6-9]

An impact is a positive or negative, direct or indirect, intended or unintended change produced by an intervention. An impact evaluation is a systematic and empirical investigation of the effects of an intervention; it assesses to what extent the outcomes experienced by affected individuals were caused by the intervention in question, and what can be attributed to other factors such as other interventions, socioeconomic trends, and political or environmental conditions. Evaluations can be categorised as formative or summative (table 1).

Approaches such as the Plan, Do, Study, Act cycle[11], which is part of the Model for Improvement, a commonly used tool to test and understand small changes in quality improvement work[12] may be used to undertake formative evaluation.

With either type of evaluation, it is important to be realistic about how long it will take to see the intended effects. Assessment that takes place too soon risks incorrectly concluding that there was no impact. This might lead stakeholders to question the value of the intervention, when later assessment might have shown a different picture. For example, in a small case study of cost savings from proactively managing high risk patients, the costs of healthcare for the

Correspondence to G Clarke Geraldine.clarke@health.org.uk

Open Access: Reuse allowed Subscribe: http://www.bmj.com/subscribe

BMJ: first published as 10.1136/bmj.l2239 on 20 June 2019. Downloaded from http://www.bmj.com/ on 8 October 2022 by guest. Protected by copyright.

eligible intervention population initially increased compared with the comparison population, but after six months were consistently lower.[14]

This article focuses on impact evaluation, but this can only ever address a fraction of questions.[15] Much more can be accomplished if it is supplemented with other qualitative and quantitative methods, including process evaluation. This provides context, assesses how the intervention was implemented, identifies any emerging unintended pathways, and is important for understanding what happened in practice and for identifying areas for improvement.[16] The economic evaluation of healthcare interventions is also important for healthcare decision making, especially with ongoing financial pressures on health services.[17]

What are the right evaluation questions?

An effective impact evaluation begins with the formulation of one or more clear questions driven by the purpose of the evaluation and what you and your stakeholders want to learn. For example, *"What is the impact of case management on patients' experience of care?"*

Formulate your evaluation questions using your understanding of the idea behind your intervention, the implementation challenges, and your knowledge of what data are available to measure outcomes. Review your theory of change or logic model[21 22] to understand what inputs and activities were planned, and what outcomes were expected and when. Once you have understood the intended causal pathway, consider the practical aspects of implementation, which include the barriers to change, unexpected changes by recipients or providers, and other influences not previously accounted for. Patient and public involvement (PPI) in setting the right question is strongly recommended for additional insights and meaningful results. For example, if evaluating the impact of case management, you could engage patients to understand what outcomes matter most to them. Healthcare leaders may emphasise metrics such as emergency admissions, but other aspects such as the experience of care might matter more to patients.[5 23]

What methods can be used to perform an impact evaluation?

Randomised control designs, where individuals are randomly selected to receive either an intervention or a control treatment, are often referred to as the "gold standard" of causal impact evaluation.[24] In large enough samples, the process of randomisation ensures a balance in observed and unobserved characteristics between treatment and control groups. However, while often suitable for assessing, for example, the safety and efficacy of medicines, these designs may be impractical, unethical, or irrelevant when assessing the impact of complex changes to health service delivery.

Observational studies are an alternative approach to estimate causal effects. They use the natural, or unplanned, variation in a population in relation to the exposure to an intervention, or the factors that affect its outcomes, to remove the consequences of a non-randomised selection process.[25] The idea is to mimic a randomised control design by ensuring treated and control groups are equivalent—at least in terms of *observed* characteristics. This can be achieved using a variety of well documented methods, including regression control and matching,[26] eg, propensity scoring[27] or genetic matching.[28] If the matching is successful at producing such groups, and there are also no differences in *unobserved* characteristics, then it can be assumed that the control group outcomes are representative of those that the treated group would have experienced if nothing had changed, ie, the counterfactual. For example, an evaluation of alternative elective surgical interventions for primary total hip replacement on osteoarthritis patients in England and Wales used genetic matching to compare patients across three different prosthesis groups, and reported that the most prevalent type of hip replacement was the least cost effective.[29]

Assessing similarity is only possible in relation to *observed* characteristics, and matching can result in biased estimates if the groups differ in relation to *unobserved* variables that are predictive of the outcome (confounders). It is rarely possible to eliminate this possibility of bias when conducting observational studies, meaning that the interpretation of the findings must always be sensitive to the possibility that the differences in outcomes were caused by a factor other than the intervention. Methods that can help when selection is on unobserved characteristics include difference-in-difference,[30] regression discontinuity,[31] instrumental variables,[18] or synthetic controls.[32] Table 2 gives a summary of selected observational study designs.

Observational studies are often referred to as natural (for natural or unplanned interventions), or quasi (for planned or intentional interventions) experiments. Natural experiments are discussed to evaluate population health interventions.[41]

What's wrong with a simple before-and-after study?

Before-and-after studies compare changes in outcomes for the *same* group of patients at a single time point before and after receiving an intervention without reference to a control group. These differ from interrupted time series studies, which compare changes in outcomes for *successive* groups of patients before and after receiving an intervention (the interruption).

Before-and-after studies are useful when it is not possible to include an unexposed control group, or for hypothesis generation. However, they are inherently susceptible to bias since changes observed may simply reflect regression to the mean (any changes in outcomes that might occur naturally in the absence of the intervention), or influences or secular trends unrelated to the intervention, eg, changes in the economic or political environment, or a heightened public awareness of issues.

For example, a before-and-after study of the impact of a care coordination service for older people tracked the hospital utilisation of the same patients before and after they were accepted into the service. They found that the service resulted in savings in hospital bed days and attendances at the emergency department.[42] Reduced hospital utilisation could have reflected regression to the mean here rather than the effects of the intervention; for example, a patient could have had a specific health crisis before being invited to join the service and then reverted back to their previous state of health and hospital utilisation for reasons unconnected with the care coordination service.

Various tools are available to evaluate the risk of bias in non-randomised designs due to confounding and other potential biases.[43 44]

Where can I find suitable routine data?

Healthcare systems generate vast amounts of data as part of their routine operation. These datasets are often designed to

BMJ: first published as 10.1136/bmj.l2239 on 20 June 2019. Downloaded from http://www.bmj.com/ on 8 October 2022 by guest. Protected by copyright.

support direct care, and for administrative purposes, rather than for research, and use of routinely collected data for evaluating changes in health service delivery is not without pitfalls. For example, any variation observed between geographical regions, providers, and sometimes individual clinicians may reflect real and important variations in the actual healthcare quality provided, but can also result from differences in measurement.[45] However, routine data can be a rich source of information on a large group of patients with different conditions across different geographical regions. Often, data have been collected for many years, enabling construction of individual patient histories describing healthcare utilisation, diagnoses, comorbidities, prescription of medication, and other treatments.

Some of these data are collected centrally, across a wider system, and routinely shared for research and evaluation purposes, eg, secondary care data in England (Hospital Episode Statistics), or Medicare Claims data in the United States. Other sources, such as primary care data, are often collected at a more local level, but can be accessed through, or on behalf of, healthcare commissioners, provided the right information governance arrangements are in place. Pseudonymised records, where any identifying information is removed or replaced by an artificial identifier, are often used to support evaluation while maintaining patient confidentiality. See table 3 for commonly used routine datasets available in England.

Healthcare records can often be linked across different sources as a single patient identifier is commonly used across a healthcare system, eg, the use of an NHS number in the UK. Using a common pseudonym across different data sources can support linkage of pseudonymised records. Linking into publicly available sources of administrative data and surveys can further enrich healthcare records. Commonly used administrative data available for UK populations include measures of GP practice quality and outcomes from the Quality and Outcomes Framework (QOF),[52] deprivation, rurality, and demographics from the 2011 Census,[53] and patient experience from the GP Patient Survey.[54]

Are there any additional considerations?

It is essential to consider threats to validity when designing and evaluating an impact evaluation; validity relates to whether an evaluation is measuring what it is claiming to measure. See Rothman et al[55] for further discussion.

Internal validity refers to whether the effects observed are due to the intervention and not some other confounding factor. Selection bias, which results from the way in which subjects are recruited, or from differing rates of participation due, for example, to age, gender, cultural or socioeconomic factors, is often a problem in non-randomised designs. Care must be taken to account for such biases when interpreting the results of an impact evaluation. Sensitivity analyses should be performed to provide reassurance regarding the plausibility of causal inferences.

External validity refers to the extent to which the results of a study can be generalised to other settings. Understanding the societal, economic, health system, and environmental context in which an intervention is delivered, and which makes its impact unique, is critical when interpreting the results of evaluations, and considering whether they apply to your setting.[56] Descriptions of context should be as rich as possible.

Often, the impact of an intervention is likely to vary depending on the characteristics of patients. These can be usefully explored in subgroup analyses.[57]

Clear and transparent reporting using established guidelines (eg, STROBE[58] or TREND[59])to describe the intervention, study population, assignment of treatment, and control groups, and methods used to estimate impact should be followed. Limitations arising as a result of inherent biases, or validity, should be clearly acknowledged.

Around the world, many interventions designed to improve health and healthcare are under way. An evaluation is an essential part of understanding what impact these changes are having, for whom and in what circumstances, and help inform future decisions about improvement and further roll out. There is no standard, "one size fits all" recipe for a good evaluation: it must be tailored to the project at hand. Understanding the overarching principles and standards is the first step towards a good evaluation.

Further Resources

See *The Health Foundation. Evaluation: what to consider. 2015*[60] for a list of websites, articles, webinars and other guidance on various aspects of impact evaluation, which may help locate further information for the planning, interpretation, and development of a successful impact evaluation.[5 23 55]

Education into practice

- What interventions have you designed or experienced aimed at transforming your service? Have they been evaluated?
- What types of routine data are collected about the care you deliver? Do you know how to access them and use them to evaluate care delivery?
- What resources are available to you to support impact evaluations for interventions?

Contributors GMC, SC, ATW and AS designed the structure of the report. GMC wrote the first draft of the manuscript. SC wrote table 2. ATW wrote table 3. AS and GMC critically revised the manuscript for important intellectual content. All authors approved the final version of the manuscript.

Competing interests We have read and understood BMJ policy on declaration of interests. All authors work in the Improvements Analytics Unit, a joint project between NHS England and the Health Foundation, which provided support for work reported in references of this report.[13 37 60]

Provenance and peer review: This article is part of a series commissioned by The BMJ based on ideas generated by a joint editorial group with members from the Health Foundation and The BMJ, including a patient/carer. The BMJ retained full editorial control over external peer review, editing, and publication. Open access fees and The BMJ's quality improvement editor post are funded by the Health Foundation.

Patient and/or members of the public were not involved in the creation of this article.

1 Davis K, Guterman S, Collins S, Stremikis G, Rustgi S, Nuzum R. *Starting on the path to a high performance health system: Analysis of the payment and system reform provisions in the Patient Protection and Affordable Care Act of.* The Commonwealth Fund, 2010, https://www.commonwealthfund.org/publications/fund-reports/2010/sep/starting-path-high-performance-health-system-analysis-payment.
2 NHS NHS Long Term Plan. 2019. https://www.england.nhs.uk/long-term-plan/
3 Djulbegovic B. A framework to bridge the gaps between evidence-based medicine, health outcomes, and improvement and implementation science. *J Oncol Pract* 2014;10:200-2. 10.1200/JOP.2013.001364 24839282
4 Bickerdike L, Booth A, Wilson PM, etal . Social prescribing: less rhetoric and more reality. A systematic review of the evidence. *BMJ Open* 2017;7:e013384.
5 Campbell M, Fitzpatrick R, Haines A, etal . Framework for design and evaluation of complex interventions to improve health. *BMJ* 2000;321:694-6. 10.1136/bmj.321.7262.694 10987780
6 Rickles D. Causality in complex interventions. *Med Health Care Philos* 2009;12:77-90. 10.1007/s11019-008-9140-4 18465202
7 Hawe P. Lessons from complex interventions to improve health. *Annu Rev Public Health* 2015;36:307-23. 10.1146/annurev-publhealth-031912-114421 25581153
8 Greenhalgh T, Papoutsi C. Studying complexity in health services research: desperately seeking an overdue paradigm shift. *BMC Med* 2018;16:95. 10.1186/s12916-018-1089-4 29921272
9 Pawson R, Tilley N. *Realistic evaluation.* Sage, 1997.

BMJ: first published as 10.1136/bmj.l2239 on 20 June 2019. Downloaded from http://www.bmj.com/ on 8 October 2022 by guest. Protected by copyright.

10 Smith J, Wistow G. (Nuffield Trust comment) Learning from an intrepid pioneer: integrated care in North West London. https://www.nuffieldtrust.org.uk/news-item/learning-from-an-intrepid-pioneer-integrated-care-in-north-west-london
11 Improvement NHS. *Plan, Do, Study, Act (PDSA) cycles and the model for improvement.* Handb Qual Serv Improv Tools, 2010.
12 Academy of Medical Royal Colleges. Quality Improvement—training for better outcomes. 2016. https://www.aomrc.org.uk/wp-content/uploads/2016/06/Quality_improvement_key_findings_140316-2.pdf
13 Lloyd T, Wolters A, Steventon A. The impact of providing enhanced support for care home residents in Rushcliffe. 2017. http://www.health.org.uk/sites/health/files/IAURushcliffe.pdf
14 Ferris TG, Weil E, Meyer GS, Neagle M, Heffernan JL, Torchiana DF. Cost savings from managing high-risk patients. In: Yong PL, Saunders RS, Olsen LA, editors. The healthcare imperative: lowering costs and improving outcomes: workshop series summary.*Nat Acad Press (US)* ; 2010:301. https://www.ncbi.nlm.nih.gov/books/NBK53910/
15 Greenhalgh T, Papoutsi C. Studying complexity in health services research: desperately seeking an overdue paradigm shift. *BMC Med* 2018;16:4-9.
16 Moore GF, Audrey S, Barker M, etal . Process evaluation of complex interventions: Medical Research Council guidance. *BMJ* 2015;350:h1258. 10.1136/bmj.h1258 25791983
17 Drummond M, Weatherly H, Ferguson B. Economic evaluation of health interventions. *BMJ* 2008;337:a1204. 10.1136/bmj.a1204 18824485
18 Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. *Stat Med* 2014;33:2297-340. 10.1002/sim.6128 24599889
19 Lorch SA, Baiocchi M, Ahlberg CE, Small DS. The differential impact of delivery hospital on the outcomes of premature infants. *Pediatrics* 2012;130:270-8. 10.1542/peds.2011-2820 22778301
20 Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Instrumental variables: application and limitations. *Epidemiology* 2006;17:260-7. 10.1097/01.ede.0000215160.88317.cb 16617274
21 Center for Theory of Change. http://www.theoryofchange.org
22 Davidoff F, Dixon-Woods M, Leviton L, Michie S. Demystifying theory and its use in improvement. *BMJ Qual Saf* 2015;24:228-38. 10.1136/bmjqs-2014-003627 25616279
23 Gertler PJ, Martinez S, Premand P, Rawlings LB, Vermeersch CMJ. Impact evaluation in practice. The World Bank Publications. 2017. https://siteresources.worldbank.org/EXTHDOFFICE/Resources/5485726-1295455628620/Impact_Evaluation_in_Practice.pdf
24 Cochrane A. Effectiveness and efficiency: random reflections on health services. London; 1972. https://www.nuffieldtrust.org.uk/research/effectiveness-and-efficiency-random-reflections-on-health-services
25 Portela MC, Pronovost PJ, Woodcock T, Carter P, Dixon-Woods M. How to study improvement interventions: a brief overview of possible study types. *Postgrad Med J* 2015;91:343-54. 10.1136/postgradmedj-2014-003620rep 26045562
26 Stuart EA. Matching methods for causal inference: A review and a look forward. *Stat Sci* 2010;25:1-21. 10.1214/09-STS313 20871802
27 Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res* 2011;46:399-424. 10.1080/00273171.2011.568786 21818162
28 Diamond A, Sekhon JS. Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *Rev Econ Stat* 2013;95:932-4510.1162/REST_a_00318.
29 Pennington M, Grieve R, Sekhon JS, Gregg P, Black N, van der Meulen JH. Cemented, cementless, and hybrid prostheses for total hip replacement: cost effectiveness analysis. *BMJ* 2013;346:f1026.
30 Wing C, Simon K, Bello-Gomez RA. Designing difference in difference studies: best practices for public health policy research. *Annu Rev Public Health* 2018;39:453-69. 10.1146/annurev-publhealth-040617-013507 29328877
31 Venkataramani AS, Bor J, Jena AB. Regression discontinuity designs in healthcare research. *BMJ* 2016;352:i1216. 10.1136/bmj.i1216 26977086
32 Abadie A, Gardeazabal J. The economic costs of conflict: a case study of the Basque country. *Am Econ Rev* 2003;93:113-3210.1257/000282803321455188.
33 Stuart EA. Matching methods for causal inference: A review and a look forward. *Stat Sci* 2010;25:1-21. 10.1214/09-STS313 20871802
34 McNamee R. Regression modelling and other methods to control confounding. *Occup Environ Med* 2005;62:500-6, 472. 10.1136/oem.2002.001115 15961628
35 Ho DE, Imai K, King G, Stuart EA. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Polit Anal* 2007;15:199-23610.1093/pan/mpl013.
36 Sutton M, Nikolova S, Boaden R, Lester H, McDonald R, Roland M. Reduced mortality with hospital pay for performance in England. *N Engl J Med* 2012;367:1821-8. 10.1056/NEJMsa1114951 23134382
37 Stephen O, Wolters A, Steventon A. Briefing: The impact of redesigning urgent and emergency care in Northumberland 2017. https://www.health.org.uk/sites/health/files/IAUNorthumberland.pdf
38 Geneletti S, O'Keeffe AG, Sharples LD, Richardson S, Baio G. Bayesian regression discontinuity designs: incorporating clinical knowledge in the causal analysis of primary care data. *Stat Med* 2015;34:2334-52. 10.1002/sim.6486 25809691
39 Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial.*Int J Epidemiol* 2016, 46:348-55.
40 Donegan K, Fox N, Black N, Livingston G, Banerjee S, Burns A. Trends in diagnosis and treatment for people with dementia in the UK from 2005 to 2015: a longitudinal retrospective cohort study. *Lancet Public Health* 2017;2667:1-8.
41 Craig P, Cooper C, Gunnell D, etal . Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. *J Epidemiol Community Health* 2012;66:1182-6. 10.1136/jech-2011-200375 22577181
42 Mayhew L. On the effectiveness of care co-ordination services aimed at preventing hospital admissions and emergency attendances. *Health Care Manag Sci* 2009;12:269-84. 10.1007/s10729-008-9092-5 19739360
43 Sterne JA, Hernán MA, Reeves BC, etal . ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919. 10.1136/bmj.i4919 27733354
44 Chen YF, Hemming K, Stevens AJ, Lilford RJ. Secular trends and evaluation of complex interventions: the rising tide phenomenon. *BMJ Qual Saf* 2016;25:303-10. 10.1136/bmjqs-2015-004372 26442789
45 Powell AE, Davies HT, Thomson RG. Using routine comparative data to assess the quality of health care: understanding and avoiding common pitfalls. *Qual Saf Health Care* 2003;12:122-8. 10.1136/qhc.12.2.122 12679509
46 NHS Digital. Hospital Episode Statistics. https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics
47 NHS Digital. Data Access Request Service (DARS) https://digital.nhs.uk/services/data-access-request-service-dars
48 NHS Digital. Secondary Uses Service (SUS) https://digital.nhs.uk/services/secondary-uses-service-sus
49 Medicines and Healthcare Regulatory Agency and National Institute for Health Research (NIHR). Clinical Practice Research Datalink(CPRD). https://www.cprd.com
50 Office for National Statistics. Deaths. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths
51 NHS Digital. Mental Health Services Data Set. https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-sets/mental-health-services-data-set
52 NHS Digital. Quality Outcomes Framework (QOF) https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/general-practice-data-hub/quality-outcomes-framework-qof
53 Office for National Statistics. 2011 Census. https://www.ons.gov.uk/census/2011census
54 NHS England. GP Patient Survey (GPPS). https://www.gp-patient.co.uk/
55 Rothman KJ, Greenland S, Lash T. *Modern Epidemiology*. Lippincott Williams & Williams, 2005.
56 Minary L, Alla F, Cambon L, Kivits J, Potvin L. Addressing complexity in population health intervention research: the context/intervention interface. *J Epidemiol Community Health* 2018;72:319-23.29321174
57 Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *BMJ* 2010;340:c117.
58 von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007;147:573-7.
59 Des Jarlais DC, Lyles C, Crepaz N. the TREND. Improving the reporting quality of nonrandomized evaluations: the TREND statement. *Am J Public Health* 2004;94:361-6.
60 The Health Foundation. Evaluation: what to consider. 2015. https://www.health.org.uk/publications/evaluation-what-to-consider

Published by the BMJ Publishing Group Limited. For permission to use (where not already granted under a licence) please go to http://group.bmj.com/group/rights-licensing/permissionsThis is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: http://creativecommons.org/licenses/by/4.0/.

BMJ: first published as 10.1136/bmj.l2239 on 20 June 2019. Downloaded from http://www.bmj.com/ on 8 October 2022 by guest. Protected by copyright.

Tables

Table 1| **Impact evaluations**

Formative	Summative	Examples
Conducted during the development or implementation of an intervention	Conducted after the intervention's completion, or at the end of a programme cycle	A formative evaluation of the Whole Systems Integrated Care (WSIC) programme, aimed at integrating health and social care in London, found that difficulties in establishing data sharing and information governance, and differences in professional culture were hampering efforts to implement change[10]
Aims to fine tune or reorient the intervention	Aims to render judgment, or make decisions about the future of the intervention	A summative impact evaluation of an NHS new care model vanguard initiative found that care home residents in Nottinghamshire who received enhanced support had substantially fewer attendances at emergency departments and fewer emergency admissions than a matched control group.[13] This evidence supported the decision by the NHS to roll out the Enhanced Health in Care Homes Model across the country.[2]

BMJ: first published as 10.1136/bmj.l2239 on 20 June 2019. Downloaded from http://www.bmj.com/ on 8 October 2022 by guest. Protected by copyright.

Table 2| Observational study designs for quantitative impact evaluation

Method	Strengths and limitations
Matching[33] Aims to find a subset of control group units (eg, individuals or hospitals) with similar characteristics to the intervention group units in the pre-intervention period. For example, impact of enhanced support in care homes in Rushcliffe, Nottinghamshire[13]	Can be combined with other methods, eg, difference-in-differences and regression. Enables straightforward comparison between intervention and control groups. Methods include propensity score matching and genetic matching
Regression control[34] Refers to use of regression techniques to estimate association between an intervention and an outcome while holding the value of the other variables constant, thus adjusting for these variables	Can be beneficial to pre-process the data using matching in addition to regression control. This reduces the dependence of the estimated treatment effect on how the regression models are specified[35]
Difference-in-differences (DiD)[30] Compares outcomes before and after an intervention in intervention and control group units. Controls for the effects of unobserved confounders that do not vary over time, eg, impact of hospital pay for performance on mortality in England[36]	Simple to implement and intuitive to interpret. Depends on the assumption that there are no unobserved differences between the intervention and control groups that vary over time, also referred to as the "parallel trends" assumption
Synthetic controls[32] Typically used when an intervention affects a whole population (eg, region or hospital) for whom a well matched control group comprising whole control units is not available. Builds a "synthetic" control from a weighted average of the control group units, eg, impact of redesigning urgent and emergency care in Northumberland[37]	Allows for unobserved differences between the intervention and control groups to vary over time. The uncertainty of effect estimates is hard to quantify. Produces biased estimates over short pre-intervention periods
Regression discontinuity design[31] Uses quasi-random variations in intervention exposure, eg, when patients are assigned to comparator groups depending on a threshold. Outcomes of patients just below the threshold are compared with those just above, eg, impact of statins on cholesterol by exploiting differences in statin prescribing[38]	There is usually a strong basis for assuming that patients close to either side of the threshold are similar. Because the method only uses data for patients near the threshold, the results might not be generalisable
Interrupted time-series[39] Compares outcomes at multiple time points before and after an intervention (interruption) is implemented to determine whether the intervention has an effect that is statistically significantly greater than the underlying trend, eg, to examine the trends in diagnosis for people with dementia in the UK[40]	Ensures limited impact of selection bias and confounding as a result of population differences but does not generally control for confounding as a result of other interventions or events occurring at the same time as the intervention
Instrumental variables[18] An instrumental variable is a variable that affects the outcome solely through the effect on whether the patient receives the treatment. An instrumental variable can be used to counteract issues of measurement error and unobserved confounders, eg, used to assess delivery of premature babies in dedicated *v* hospital intensive care units[19]	Explicitly addresses unmeasured confounding but conceptually difficult and easily misused. Identification of instrumental variables is not straightforward. Estimates are imprecise (large standard error), biased when sample size is small, and can be biased in large samples if assumptions are even slightly violated[20]

BMJ: first published as 10.1136/bmj.l2239 on 20 June 2019. Downloaded from http://www.bmj.com/ on 8 October 2022 by guest. Protected by copyright.

Table 3| **Commonly used routine datasets available in the NHS in England**

Dataset	Dissemination and alternatives
Hospital episode statistics (HES).[46] HES is a database containing details of all admissions, accident and emergency attendances, and outpatient appointments at NHS England hospitals and NHS England funded treatment centres. Information captured includes clinical information about diagnoses and operations, patient demographics, geographical information, and administrative information such as the data and method of admissions and discharge	HES is available through the Data Access Request Service (DARS),[47] a service provided by NHS Digital. Commissioners, providers in the NHS, and analytics teams working on their behalf, can also access hospital data directly via the Secondary Use Service (SUS).[48] These data are very similar to HES, processed by NHS Digital, and are available for non-clinical uses, including research and planning health services
Primary care data is collected by general practices. Although there is no national standard on how primary care data should be collected and/or reported, there are a limited number of commonly used software providers to record these data. Information captured includes clinical information about diagnoses, treatment, and prescriptions, patient demographics, geographical information, and administrative information on booking and attendance of appointments, and whether appointments relate to a telephone consultation, an in-practice appointment, or a home visit	Commissioners, and analytics teams working on their behalf, can work with an intermediary service called Data Service for Commissioning Regional Office to request access to anonymised patient level general practice data (possibly linked to SUS, described above) for the purpose of risk stratification, invoice validation, and to support commissioning. Anonymised UK primary care records for a representative sample of the population are available for public health research through, for instance, the Clinical Practice Research Datalink.[49]
Mortality data[50] The Office for National Statistics (ONS) maintains a dataset of all registered deaths in England. These data can be linked to routine health data to record deaths that occur outside of hospital	ONS mortality data are routinely processed by NHS Digital, and can be linked to HES data. These data can be requested through the DARS service. When deaths occur in hospital this is typically recorded as part of discharge information
The Mental Health Services Data Set (MHSDS)[51] contains record level data about the care of children, young people, and adults who are in contact with mental health, learning disabilities, or autism spectrum disorder services. These data cover data from April 2016	Like HES, MHSDS is available through the DARS service. Mental health data from before April 2016 have been recorded in the Mental Health Minimum Dataset also disseminated through NHS Digital

BMJ: first published as 10.1136/bmj.l2239 on 20 June 2019. Downloaded from http://www.bmj.com/ on 8 October 2022 by guest. Protected by copyright.