Foy *et al. Trials* (2023) 24:537
https://doi.org/10.1186/s13063-023-07578-5

COMMENTARY **Open Access**

What is the role of randomised trials in implementation science?

Robbie Foy[1*], Noah M. Ivers[2], Jeremy M. Grimshaw[3] and Paul M. Wilson[4]

Abstract

Background There is a consistent demand for implementation science to inform global efforts to close the gap between evidence and practice. Key evaluation questions for any given implementation strategy concern the assessment and understanding of effects. Randomised trials are generally accepted as offering the most trustworthy design for establishing effectiveness but may be underused in implementation science.

Main body There is a continuing debate about the primacy of the place of randomised trials in evaluating implementation strategies, especially given the evolution of more rigorous quasi-experimental designs. Further critiques of trials for implementation science highlight that they cannot provide 'real world' evidence, address urgent and important questions, explain complex interventions nor understand contextual influences. We respond to these critiques of trials and highlight opportunities to enhance their timeliness and relevance through innovative designs, embedding within large-scale improvement programmes and harnessing routine data.

Our suggestions for optimising the conditions for randomised trials of implementation strategies include strengthening partnerships with policy-makers and clinical leaders to realise the long-term value of rigorous evaluation and accelerating ethical approvals and decluttering governance procedures for lower risk studies.

Conclusion Policy-makers and researchers should avoid prematurely discarding trial designs when evaluating implementation strategies and work to enhance the conditions for their conduct.

Keywords Cluster randomised controlled trials, Implementation science, Quality improvement

Background

Gaps between evidence and practice pervade different healthcare systems and specialties [1–7]. They are associated with poorer patient outcomes [3–5, 8] and can be resistant to change [6, 9, 10]. There is over-treatment as well as under-treatment [11, 12]. During the COVID-19 pandemic, access to medical interventions such as vaccines was inconsistently matched to population need [13].

These research-to-practice gaps represent a strategically important problem for policy-makers, healthcare systems and research funders because they limit the health, social and economic impacts of research [14]. Regular processions of policy reforms aim to tackle inappropriate variations to ensure the quality and safety of healthcare. Their effects, efficiency and reach frequently fall short of initial hopes and expectations [9, 15, 16]. Indeed, there are serious risks of wasted effort if urgent drives to improve care are insufficiently underpinned by scientific methods [17]. There is therefore a predictable and sustained need for implementation science, which aims to inform policy decisions about how best to use

*Correspondence:
Robbie Foy
r.foy@leeds.ac.uk
[1] Leeds Institute of Health Sciences, University of Leeds, Leeds, UK
[2] Women's College Hospital, Department of Family and Community Medicine, University of Toronto, Toronto, Canada
[3] Ottawa Hospital Research Institute, Ottawa, Canada
[4] Centre for Primary Care and Health Services Research, University of Manchester, Manchester, UK

resources to improve the uptake of research findings by evaluating approaches to change clinical and organisational behaviour [18]. Such approaches may include interventions such as audit and feedback (providing a summary of clinical performance over a specified period of time), computerised decision support systems and financial incentives [19]. Implementation interventions typically aim to increase uptake of clinical interventions of established effectiveness (e.g. intensification of treatment for type 2 diabetes [20]) or reduce use of low value or harmful clinical interventions (e.g. x-rays for non-specific low back pain [21]).

There are key evaluation questions of interest to patients, professionals and policy-makers for any given implementation strategy. Does it work? For which contextual features and targeted behaviours is it likely to work? Are the costs worth the benefits?

Well-conducted randomised controlled trials offer a 'fair test' of effectiveness by balancing known and unknown confounders so that differences in outcomes between comparison groups can be confidently attributed to intervention effects. Given that implementation strategies usually target organisations, cluster randomisation (e.g. of general practices or hospitals) is usually more appropriate than individual patient randomisation [22, 23]. Across health systems, there is a growing impetus to innovate, experiment and rapidly implement new solutions for health care problems—this was especially true during the COVID-19 pandemic. However, there is a continuing debate about the place of trials in the field of implementation science [24–28], especially given that many strategies to change professional and organisational behaviours can be conceptualised as complex interventions [29].

We respond to critiques of trials for implementation science and highlight opportunities to enhance their design, delivery and efficiency. We suggest that policy-making and research communities work to optimise the conditions for conducting trials.

Main text

Critiques of trials

Alternative evaluation designs offer similar protections against bias

There are natural, human tendencies to look for expected or even hoped for intervention effects, and hence reach erroneous conclusions from an evaluation. Comparisons of more and less rigorous evaluation methods suggest that the best way to show that an intervention 'works' is to use a weak non-randomised (quasi-experimental) study design without a concurrent control group [30, 31]. However, more rigorous quasi-experimental evaluation designs offer a viable alternative to trials for evaluating implementation strategies [32–35]. For example, a reanalysis of cluster randomised trials as interrupted time series found that effect estimates were largely concordant [36]. Rigorous quasi-experimental studies can make important policy contributions where randomisation was not acceptable or feasible, for example, in examining the effects of withdrawing financial incentives on adherence to primary care indicators [37].

Whilst relatively sophisticated quasi-experimental designs hold considerable promise, further experience is needed to understand their properties, strengths and limitations in much the same way that it has taken decades to develop evidence-informed criteria to judge the validity and generalisability of trials [38]. Such criteria also guide researchers to anticipate and implement established measures to reduce observational and performance biases.

The understanding of known and unknown confounders within the context of healthcare systems is relatively poor compared to, say, biological systems. Multiple contextual influences on the outcomes of implementation strategies [39] make attribution of intervention effects challenging, if not heroic, in the absence of rigorously controlled evaluations. Non-randomised designs may not be able to frame or rule out plausible rival hypotheses to any apparent intervention effect, particularly within the complex and evolving contexts of healthcare systems. The interpretation of changes over time is particularly vulnerable to system-level disruptions, the COVID-19 pandemic being an extreme case.

Confident attribution is important given that the observed effect sizes of implementation interventions can be small, even if worthwhile from population perspectives. For example, a trial of a multifaceted implementation strategy found just over a 1% absolute difference in high-risk prescribing between intervention and control general practices, which translated into a cost-effective population benefit from reduced patient harm [40].

Trials cannot provide 'real world' evidence

Clinical trials are dogged by the criticism that they recruit highly selected participants atypical of wider patient populations who may also receive above average attention, thereby limiting generalisability [41]. Such a criticism could justly be levelled at an implementation trial that recruits selected sites and delivers interventions requiring resources and skillsets not typically available to healthcare systems. Yet, most implementation trials tend to be pragmatic in several ways [42], by recruiting unselected sites, adapting available interventions, allowing flexibility of delivery and using non-intrusive data collection to assess outcomes [40]. For example, Elouafkaoui et al. randomly allocated all 795 National Health Service

general dental practices in Scotland to receive or not to receive individualised feedback, derived from routinely collected data, and found that feedback reduced the antibiotic prescribing rate by 5.7% [43]. In contrast, less rigorous effectiveness evaluations may include relatively small numbers of volunteer sites with limited generalisability [22, 44].

Trials cannot address urgent and important questions

Almost every healthcare problem, such as avoidable cancer deaths and strokes or maximising COVID vaccination coverage, demands urgent solutions. The history of medicine is littered with obsolete clinical recommendations based on partial evidence and assumptions, subsequently overturned by rigorous studies [45–49]. Similarly, major and costly initiatives have ended or been threatened because of continuing uncertainty over their benefits because trials were considered unacceptable and unfeasible [50–52]; given the cyclical nature of policy reforms [53], it is likely that similar initiatives will emerge again and be under-evaluated again.

Alternative evaluation designs offer the attraction of faster turnaround times than trials, with shorter planning and follow-up periods. However, time series designs depend upon stable data being available over lengthy periods. Some of the long timelines associated with implementation trials are related to burdensome research regulation and management [54], a limitation of the wider system rather than trials per se.

Well-conducted trials can overturn conventional wisdom whilst positive quasi-experimental studies may fail to convince. A trial of a multidisciplinary intervention targeting evidence-based management of fever, hyperglycaemia and swallowing dysfunction in acute stroke units found significant reductions in patient deaths or dependency at 90 days [55] and in deaths after 4 years [56]. The clinical benefits were markedly greater than those observed for other interventions, such as stroke units or thrombolysis. It might otherwise have been more difficult to convince sceptics of the value of the multidisciplinary intervention.

Trials shed little light on complex interventions

There are challenges in evaluating complex interventions, which contain several interacting components and target one or more behaviours, target more than one group or organisational level, result in different numbers and types of outcome and permit degrees of tailoring or flexibility [57]. Some of these interventions, such as digital health records and service configurations, may evolve over time and become outmoded before any evaluation is completed.

It has been suggested that whilst randomised trials may be appropriate for clinical interventions, 'service innovations are even more complex, and this complexity needs to be embraced, not eliminated' [27]. Hence, mixed method evaluations incorporating quasi-experimental designs are better suited to evaluating large-scale service reconfigurations, such as hyperacute stroke care [58]. However, alternative methods of evaluating complex interventions may have their own pitfalls. For example, multiple case studies are unlikely to include a robust assessment of whether any intervention works and unlikely to have a large and representative enough sample to allow generalizable conclusions.

Some complex interventions are not ready for trial evaluation. The UK Medical Research Council framework for the development and evaluation of complex interventions recommends the use of multiple research methods, with scope for embedding trials within a broader programme of studies which can also contribute to understanding mechanisms of change [29]. The framework further recognises the need for a sound theoretical understanding of causality (e.g. within a logic model) and hence the definition of prototypical elements followed by feasibility testing in context to help decide when an evolving intervention is stable enough for trial evaluation [59].

Trials shed little light on contextual influences

Contextual factors can have major influences on intervention effects. A criticism of trials is that by controlling for and 'eliminating' the influences of contextual factors, trials cannot provide information about their impacts on change [60]. This criticism may apply to any effectiveness evaluation. Trials can help understand contextual influences in three ways. First, they provide an opportunity to not only look at the mean effect but to explore whether contextual variations matter, with less concern about unknown confounders. For example, a trial in general practice demonstrated that antimicrobial stewardship, comprising a webinar, monthly feedback reports and electronic decision support, was effective for adults but not children, suggesting the need for an alternative approach for a different patient population [61]. Second, qualitative and quantitative process evaluations, ideally conducted in parallel to trials, also generate insights into contextual influences [62]. A process evaluation indicated that a multifaceted strategy to improve induced abortion care was ineffective because gynaecology teams were already highly motivated to follow best practice guidance but hindered by organisational constraints [63]. Third, comparing findings of similar interventions for different targeted behaviours or across different settings allows indirect comparisons of contextual modifiers, especially

via systematic reviews. A meta-analysis of 122 trials of computerised clinical decision support systems found that low baseline adherence and paediatric settings were associated with significantly larger absolute improvements in care [64]. Thus, pursuing a rigorous answer to the question of 'whether' an implementation strategy worked is not mutually exclusive to—and may in fact facilitate—elaborations of theory regarding 'where', 'how', and 'why'.

Innovations and opportunities

Identifying and prioritising 'best bet' interventions

Implementation interventions typically have several components but conducting multiple trials of every permutation can be wasteful. For example, varying only five elements of audit and feedback (e.g. differing frequencies of feedback) produces 288 combinations—not allowing for replication of studies or the addition of other interventions, such as educational meetings or outreach visits [65]. Some trial designs allow for adaptation of intervention components or assignment to interventions as evaluations proceed.

The Multiphase Optimization Strategy (MOST) offers a methodological approach for building, optimising and evaluating multicomponent interventions. MOST comprises three steps: preparation, laying the groundwork for optimisation by conceptualising and piloting components; optimisation, conducting trials to identify the most promising single or combined intervention components; and evaluation, a definitive randomised trial to assess intervention effectiveness [66]. Modelling experiments can identify and prioritise the most promising 'active ingredients' for further study [67]. These experiments can be conducted virtually (e.g. online) with targeted participants using proxy outcomes (e.g. behavioural intentions) [68].

The Sequential Multiple Assignment Randomized Trial (SMART) allows identification of the best tailoring variables and uses decision rules for adaptive interventions based upon early findings. It is especially suited for building time-varying adaptive interventions. It has been used to tailor the intensity of an intervention to improve uptake of a re-engagement programme for patients with serious mental illness according to site characteristics and initial responses to interventions [69].

The stepped wedge design offers a solution where there is uncertainty, but randomisation to a non-intervention control is unacceptable. It entails introducing an intervention to groups of clusters in a random order. There are no 'losers' because all sites eventually receive the intervention. A stepped wedge trial demonstrated that an intervention comprising professional education, informatics to facilitate review and financial incentives reduced high-risk prescribing in general practices [70]. Stepped wedge trials can be complex to conduct [71] and their analysis fraught with pitfalls [72, 73]. One assumption, that the intervention does no harm, may not hold; a stepped wedge trial of predictive risk stratification to identify and manage patients at higher risk of emergency hospital admissions found that it increased emergency attendances, hospitalisation and costs without benefiting patients [74].

Implementation laboratories

Trials offer opportunities to optimise the effectiveness of existing implementation interventions, in much the same way that clinical research has continually pushed marginal gains in the effective management of conditions such as cancer or stroke. Yet, establishing the infrastructure for each new trial can be costly and time-consuming. There are opportunities for implementation researchers to learn from and adapt methodologies from clinical fields, such as oncology; innovations such as 'master protocols' are based upon a single overarching design to evaluate multiple hypotheses with the goal of improving efficiency and standardising the development and evaluation of different interventions [75].

Large-scale programmes offer opportunities for embedded trials. The PRevention of Cerebral Palsy in Pre-Term labour (PReCePT) programme aimed to reduce cerebral palsy by promoting the use of magnesium sulphate in pregnant women at risk of premature delivery in England. The programme included a nested trial comparing two approaches to quality improvement [76].

The next evolutionary step is to create a learning health system which makes small, incremental changes supported by tightly focused evaluations, and thereby cumulatively improves patient care whilst developing the underpinning evidence base. Such 'radical incrementalism' offers a potentially cost-effective if underutilised approach to embedding learning within large scale improvement programmes [77]. It has already been used in public policy and in business [78]; *Amazon* and *eBay* randomise potential customers to different presentations of their products online to understand what drives purchases. It is also applicable to healthcare. For example, within a national clinical audit programme, is feeding back data on performance indicating an organisation's position against the top 10% more likely to stimulate improvement than showing its position against median performance? Does adding quality improvement facilitation to standard feedback have effects over standard feedback alone? Implementation laboratories entail embedding a sequential programme of head-to-head trials testing different versions of interventions within an established improvement initiative [79, 80]. Those

Table 1 Suggestions for optimising the conditions for randomised trials of implementation strategies

Anticipating the need to design and conduct trials should quasi-experimental evaluations show promising results insufficient to shift equipoise
Building partnerships with policy-makers and clinical leaders and promoting the long-term value of rigorous evaluation
Funding programmes of trials to evaluate progressive incremental changes to implementation strategies
Embedding the design and delivery of trials within large scale improvement programmes
Accelerating ethical approvals and decluttering governance procedures for lower risk studies
Reducing the burden and intrusiveness of outcome assessment by using routinely collected data

versions identified as more effective than the current standard become the new standard whilst those which are not more effective are discarded. The UK National Clinical Audit of Blood Transfusions collaborated in trials comparing different ways of presenting content and supporting delivery of feedback reports to hospitals [81]. Similar opportunities apply to other frequently-used implementation strategies, such as clinical decision support systems or educational programmes.

Implementation laboratories several further advantages. First, they can reduce research waste [82], such as the failure to build upon empirical findings in developing and evaluating feedback interventions [83]. Second, cluster randomised trials typically require larger numbers of patients than individually randomised trials to account for lack of independence at the cluster level. Increasing the number of sites generally buys greater statistical efficiency than increasing the number of patients. Embedding trials within an existing network or major improvement initiative facilitates recruitment, data collection and helps ensure 'real world' generalisability, building on the advantages of registry-based trials [84]. Third, comparing and integrating findings from different implementation laboratories through a 'meta-laboratory' allows learning about important contextual effect modifiers and mediators.

Harnessing routinely collected data

The collection and analysis of project–specific data is expensive and limits the volume and duration of data collection. Routinely collected data can be applied to develop quality indicators [85], analyse variations in care [12] and assess outcomes in trials of implementation strategies [86, 87]. Routine 'big' datasets offer opportunities to improve research efficiency, improve internal validity via non-intrusive data collection (and reduce risk of Hawthorne effects) and enhance generalisability and reach through participation of unselected healthcare provider and patient populations. For example, a trial of practice facilitation to support family physicians to engage with their patients around COVID-19 vaccinations used existing administrative databases to identify the practices with greatest need and allocated them to the offer of support or usual care [88].

Trials using routinely collected data may also be able to achieve relatively large sample sizes, hence bolstering statistical power to detect modest effect sizes and explore effect modifiers. However, larger samples may be needed to compensate for additional 'noise' from using data not originally intended for research. It is important to ensure that any routinely available data are a good fit for the outcomes of interest or based upon reasonable assumptions about relevance. Using unplanned hospital readmission rates in evaluating interventions to improve the process of hospital discharge to patient homes assumes that most such readmissions are driven by adverse events and would rather be avoided by patients and healthcare systems. Just as innovative 'platform trials' have been crucial to guide the clinical treatment of COVID-19 [89], custom-built registries used by national clinical audit programmes offer platforms for implementation trials [90]. They also provide a means for monitoring subsequent uptake and population impact of implementation strategies beyond trial lifetimes [91].

Optimising conditions for trials

Increasing burdens of research regulation and management may have unwittingly conspired to undermine the feasibility and timeliness of trials [54]. Experience of the COVID-19 pandemic has demonstrated that approvals and governance procedures can be streamlined with sufficient will [92, 93]. There are calls to make the conduct of trials for drug development easier, faster and cheaper [94]. There is an equally strong case for similar actions around lower-risk research which aims to accelerate the uptake of evidence-based practice. Table 1 suggests some avenues to explore in optimising the conditions for the conduct of implementation trials.

Conclusion

Trials generally offer known protection against threats to internal validity, chiefly selection bias, in the evaluation of implementation strategies. Their findings are less dependent on skilled and nuanced interpretation compared to other study designs. Pragmatic trials can provide real world evidence in addressing important implementation problems and improve understanding of both complex interventions and contextual influences. There are

opportunities to advance implementation science and its impact through innovative trial designs, implementation laboratories and the use of routine data. We encourage researchers, funders and policy-makers to consider when randomised evaluations would be feasible and preferable and work to optimise conditions for their conduct.

Authors' contributions
All authors have been involved in the conception and drafting of the paper and have approved the final version. RF led the drafting of the paper and is guarantor.

Funding
Robbie Foy is the Implementation Laboratory lead for the National Institute for Health and Care Research (NIHR) Yorkshire and Humber Applied Research Collaboration. Paul Wilson is the Implementation Science research theme lead for the NIHR Greater Manchester Applied Research Collaboration. This report is independent research funded by the NIHR Yorkshire and Humber Applied Research Collaboration (NIHR 200166) and the NIHR Greater Manchester Applied Research Collaboration. The views expressed in this publication are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.
Noah Ivers holds a Canada Research Chair in Implementation of Evidence-based Practice. Jeremy Grimshaw holds a Canada Research Chair in Health Knowledge Transfer and Uptake.

Availability of data and materials
Not applicable.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
Robbie Foy, Noah Ivers and Jeremy Grimshaw are members of the Audit and Feedback MetaLab. The MetaLab is an international collaboration to advance learning and expertise on Audit and Feedback and promotes the development of audit and feedback implementation laboratories internationally. Paul Wilson is Co-Editor in Chief of Implementation Science, and Robbie Foy, Noah Ivers and Jeremy Grimshaw are Editorial Board members.

Received: 8 May 2023 Accepted: 4 August 2023
Published online: 16 August 2023

References

1. Seddon ME, Marshall MN, Campbell SM, Roland MO. Systematic review of studies of quality of clinical care in general practice in the UK, Australia and New Zealand. Qual Health Care. 2001;10:152–8.
2. Steel N, Bachmann M, Maisey S, Shekelle P, Breeze E, Marmot M, Melzer D. Self reported receipt of care consistent with 32 quality indicators: national population survey of adults aged 50 or more in England. Bmj. 2008;337:a957.
3. Zeitlin J, Manktelow BN, Piedvache A, Cuttini M, Boyle E, van Heijst A, Gadzinowski J, Van Reempts P, Huusom L, Weber T, et al. Use of evidence based practices to improve survival without severe morbidity for very preterm infants: results from the EPICE population based cohort. Bmj. 2016;354:i2976.
4. Wright AA, Cronin A, Milne DE, Bookman MA, Burger RA, Cohn DE, Cristea MC, Griggs JJ, Keating NL, Levenback CF, et al. Use and effectiveness of intraperitoneal chemotherapy for treatment of ovarian cancer. J Clin Oncol. 2015;33(26):2841–7.
5. Chung SC, Sundstrom J, Gale CP, James S, Deanfield J, Wallentin L, Timmis A, Jernberg T, Hemingway H. Comparison of hospital variation in acute myocardial infarction care and outcome between Sweden and United Kingdom: population based cohort study using nationwide clinical registries. BMJ. 2015;351:h3913.
6. Levine DM, Linder JA, Landon BE. The quality of outpatient care delivered to adults in the United States, 2002 to 2013. JAMA Intern Med. 2016;176(12):1778–90.
7. Turner GM, Calvert M, Feltham MG, Ryan R, Fitzmaurice D, Cheng KK, Marshall T. Under-prescribing of prevention drugs and primary prevention of stroke and transient ischaemic attack in UK general practice: a retrospective analysis. PLoS Med. 2016;13(11): e1002169.
8. Bucholz EM, Butala NM, Normand SL, Wang Y, Krumholz HM. Association of guideline-based admission treatments and life expectancy after myocardial infarction in elderly medicare beneficiaries. J Am Coll Cardiol. 2016;67(20):2378–91.
9. Serumaga B, Ross-Degnan D, Avery A, Elliott R, Majumdar SR, Zhang F, Soumerai SB. Effect of pay for performance on the management and outcomes of hypertension in the United Kingdom: interrupted time series study. BMJ. 2011;342:d108.
10. Campbell SM, Roland MO, Middleton E, Reeves D. Improvements in quality of clinical care in English general practice 1998–2003: longitudinal observational study. BMJ. 2005;331(7525):1121.
11. Fleming-Dutra KE, Hersh AL, Shapiro DJ, Bartoces M, Enns EA, File TM Jr, Finkelstein JA, Gerber JS, Hyun DY, Linder JA, et al. Prevalence of inappropriate antibiotic prescriptions among US ambulatory care visits, 2010–2011. JAMA. 2016;315(17):1864–73.
12. Foy R, Leaman B, McCrorie C, Petty D, House A, Bennett M, Carder P, Faulkner S, Glidewell L, West R. Prescribed opioids in primary care: cross-sectional and longitudinal analyses of influence of patient and practice characteristics. BMJ Open. 2016;6(5):e010276.
13. Yamey G, Garcia P, Hassan F, Mao W, McDade KK, Pai M, Saha S, Schellekens P, Taylor A, Udayakumar K. It is not too late to achieve global covid-19 vaccine equity. BMJ. 2022;376:e070650.
14. Cooksey R. A review of UK health research funding. Norwich: HMSO; 2006. Accessed 4 Sept 2022.
15. Roland M. Linking physicians' pay to the quality of care – a major experiment in the United Kingdom. N Engl J Med. 2004;351:1448–54.
16. Ryan AM, Krinsky S, Kontopantelis E, Doran T. Long-term evidence for the effect of pay-for-performance in primary care on mortality in the UK: a population study. Lancet. 2016;388(10041):268–74.
17. Auerbach A, Landefeld C, Shojania K. The tension between needing to improve care and knowing how to do it. N Engl J Med. 2007;357:608–13.
18. Foy R, Eccles M, Grimshaw J. Why does primary care need more implementation research? Family Practice. 2001;18:353–5.
19. Grimshaw JM, Eccles MP, Lavis JN, Hill SJ, Squires JE. Knowledge translation of research findings. Implement Sci. 2012;7:50.
20. Furler J, O'Neal D, Speight J, Manski-Nankervis J-A, Gorelik A, Holmes-Truscott E, Ginnivan L, Young D, Best J, Patterson E, et al. Supporting insulin initiation in type 2 diabetes in primary care: results of the Stepping Up pragmatic cluster randomised controlled clinical trial. BMJ. 2017;356:j783.
21. Eccles M, Steen N, Grimshaw J, Thomas L, McNamee P, Soutter J, Wilsdon J, Matowe L, Needham G, Gilbert F, et al. Effect of audit and feedback, and reminder messages on primary-care radiology referrals: a randomised trial. Lancet. 2001;357:1406–9.
22. Eccles M, Grimshaw JM, Campbell M, Ramsay C. Research designs for studies evaluating the effectiveness of change and quality improvement strategies. Qual Saf Health Care. 2003;12:47–52.
23. Wolfenden L, Foy R, Presseau J, Grimshaw JM, Ivers NM, Powell BJ, Taljaard M, Wiggers J, Sutherland R, Nathan N, et al. Designing and undertaking randomised implementation trials: guide for researchers. BMJ. 2021;372:m3721.
24. Davies HTO (ed), Nutley S (ed), Smith PC (ed). What works? Evidence-based policy and practice in public services. Bristol: Policy Press; 2000.
25. Berwick DM. The science of improvement. JAMA. 2008;283:2275–80.
26. Oliver D. Should practical quality improvement have parity of esteem with evidence based medicine? BMJ. 2017;357:j2582.

27. Lamont T, Barber N, de Pury J, Fulop N, Garfield-Birkbeck S, Lilford R, Mear L, Raine R, Fitzpatrick R. New approaches to evaluating complex health and care systems. BMJ. 2016;352:i154.
28. Wensing M, Grol R. Knowledge translation in health: how implementation science could contribute more. BMC Med. 2019;17(1):88.
29. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, Boyd KA, Craig N, French DP, McIntosh E, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ. 2021;374:n2061.
30. Soumerai SB, McLaughlin TJ, Avorn J. Improving drug prescribing in primary care: a critical analysis of the experimental literature. Milbank Q. 1989;67:268–317.
31. Foy R, Hempel S, Rubenstein L, Suttorp M, Seelig M, Shanman R, Shekelle P. Meta-analysis: effect of interactive communication between collaborating primary care physicians and specialists. Ann Intern Med. 2010;152:247–58.
32. Kontopantelis E, Doran T, Springate DA, Buchan I, Reeves D. Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. BMJ. 2015;350:h2750.
33. Venkataramani AS, Bor J, Jena AB. Regression discontinuity designs in healthcare research. Bmj. 2016;352:i1216.
34. Barnighausen T, Tugwell P, Rottingen JA, Shemilt I, Rockers P, Geldsetzer P, Lavis J, Grimshaw J, Daniels K, Brown A, et al. Quasi-experimental study designs series-paper 4: uses and value. J Clin Epidemiol. 2017;89:21–9.
35. Barr B, Zhang X, Green M, Buchan I. A blueprint for synthetic control methodology: a causal inference tool for evaluating natural experiments in population health. BMJ. 2022;379:o2712.
36. Fretheim A, Zhang F, Ross-Degnan D, Oxman AD, Cheyne H, Foy R, Goodacre S, Herrin J, Kerse N, McKinlay RJ, et al. A reanalysis of cluster randomized trials showed interrupted time-series studies were valuable in health system evaluation. J Clin Epidemiol. 2015;68(3):324–33.
37. Morales DR, Minchin M, Kontopantelis E, Roland M, Sutton M, Guthrie B. Estimated impact from the withdrawal of primary care financial incentives on selected indicators of quality of care in Scotland: controlled interrupted time series analysis. BMJ. 2023;380:e072098.
38. Campbell MK, Elbourne DR, Altman DG. for the CG: CONSORT statement: extension to cluster randomised trials. BMJ. 2004;328:702–8.
39. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci. 2009;4:50.
40. Willis TA, Collinson M, Glidewell L, Farrin AJ, Holland M, Meads D, Hulme C, Petty D, Alderson S, Hartley S, et al. An adaptable implementation package targeting evidence-based indicators in primary care: a pragmatic cluster-randomised evaluation. PLOS Med. 2020;17(2):e1003045.
41. Treweek S, Zwarenstein M. Making trials matter: pragmatic and explanatory trials and the problem of applicability. Trials. 2009;10:37.
42. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. BMJ. 2015;350:h2147.
43. Elouafkaoui P, Young L, Newlands R, Duncan EM, Elders A, Clarkson JE. An audit and feedback intervention for reducing antibiotic prescribing in general dental practice: the RAPiD cluster randomised controlled trial. PLoS Med. 2016;13(8):e1002115.
44. Shadish WR, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalised causal inference. Boston: Houghton Mifflin Company; 2002.
45. Fleming PJ, Gilbert R, Azaz Y, Berry PJ, Rudd PT, Stewart A, Hall E. Interaction between bedding and sleeping position in the sudden infant death syndrome: a population based case-control study. BMJ. 1990;301(6743):85–9.
46. Cochrane AL. Effectiveness and efficiency: random reflections on health services. London: Nuffield Provincial Hospitals Trust; 1972.
47. Edwards P, Arango M, Balica L, Cottingham R, El-Sayed H, Farrell B, Fernandes J, Gogichaisvili T, Golden N, Hartzenberg B, et al. Final results of MRC CRASH, a randomised placebo-controlled trial of intravenous corticosteroid in adults with head injury-outcomes at 6 months. Lancet. 2005;365(9475):1957–9.
48. The AVERT Trial Collaboration group. Efficacy and safety of very early mobilisation within 24 h of stroke onset (AVERT): a randomised controlled trial. Lancet. 2015;386(9988):46–55.
49. Elshaug AG, Watt AM, Mundy L, Willis CD. Over 150 potentially low-value health care practices: an Australian study. Med J Aust. 2012;197(10):556–60.
50. Roland M, Guthrie B. Quality and Outcomes Framework: what have we learnt? BMJ. 2016;354:i4060.
51. Wilson PM, Boaden R, Harvey G. Plans to accelerate innovation in health systems are less than IDEAL. BMJ Qual Saf. 2016;25(8):572–6.
52. Westlake D, Tierney S, Wong G, Mahtani KR. Social prescribing in the NHS—is it too soon to judge its value? BMJ. 2023;380: p699.
53. Klein R. The new politics of the National Health Service. London: Longman Publishing Group; 1995.
54. Al-Shahi Salman R, Beller E, Kagan J, Hemminki E, Phillips RS, Savulescu J, Macleod M, Wisely J, Chalmers I. Increasing value and reducing waste in biomedical research regulation and management. Lancet. 2014;383(9912):176–85.
55. Middleton S, McElduff P, Ward J, Grimshaw JM, Dale S, D'Este C, Drury P, Griffiths R, Cheung NW, Quinn C, et al. Implementation of evidence-based treatment protocols to manage fever, hyperglycaemia, and swallowing dysfunction in acute stroke (QASC): a cluster randomised controlled trial. Lancet. 2011;378(9804):1699–706.
56. Middleton S, Coughlan K, Mnatzaganian G, Low Choy N, Dale S, Jammali-Blasi A, Levi C, Grimshaw JM, Ward J, Cadilhac DA, et al. Mortality reduction for fever, hyperglycemia, and swallowing nurse-initiated stroke intervention: QASC Trial (Quality in Acute Stroke Care) follow-up. Stroke. 2017;48(5):1331–6.
57. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ. 2008;337:a1655.
58. Morris S, Hunter RM, Ramsay AI, Boaden R, McKevitt C, Perry C, Pursani N, Rudd AG, Schwamm LH, Turner SJ, et al. Impact of centralising acute stroke services in English metropolitan areas on mortality and length of hospital stay: difference-in-differences analysis. BMJ. 2014;349:g4757.
59. Shojania KG. Conventional evaluations of improvement interventions: more trials or just more tribulations? BMJ Qual Saf. 2013;22(11):881–4.
60. Pawson R, Tilley N. Realistic evaluation. London: Sage; 1997.
61. Gulliford MC, Prevost AT, Charlton J, Juszczyk D, Soames J, McDermott L, Sultana K, Wright M, Fox R, Hay AD, et al. Effectiveness and safety of electronically delivered prescribing feedback and decision support on antibiotic use for respiratory illness in primary care: REDUCE cluster randomised trial. BMJ. 2019;364:l236.
62. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, Moore L, O'Cathain A, Tinati T, Wight D, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ. 2015;350: h1258.
63. Foy R, Walker A, Ramsay C, Penney G, Grimshaw J, Francis JJ. Theory-based identification of barriers to quality improvement: induced abortion care. Int J Qual Health Care. 2005;17:147–55.
64. Kwan JL, Lo L, Ferguson J, Goldberg H, Diaz-Martinez JP, Tomlinson G, Grimshaw JM, Shojania KG. Computerised clinical decision support systems and absolute improvements in care: meta-analysis of controlled clinical trials. BMJ. 2020;370:m3216.
65. The Improved Clinical Effectiveness through Behavioural Research Group. Designing theoretically-informed implementation interventions. Implement Sci. 2006;1:4.
66. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. Am J Prev Med. 2007;32(5 Suppl):S112-118.
67. Eccles MP, Francis J, Foy R, Johnston M, Bamford C, Grimshaw JM, Hughes J, Lecouturier J, Steen N, Whitty PM. Improving professional practice in the disclosure of a diagnosis of dementia: a modeling experiment to evaluate a theory-based intervention. Int J Behav Med. 2009;16(4):377–87.
68. Gude WT, van Engen-Verheul MM, van der Veer SN, de Keizer NF, Peek N. How does audit and feedback influence intentions of health professionals to improve practice? A laboratory experiment and field study in cardiac rehabilitation. BMJ Qual Saf. 2017;26(4):279–87.
69. Kilbourne AM, Almirall D, Goodrich DE, Lai Z, Abraham KM, Nord KM, Bowersox NW. Enhancing outreach for persons with serious mental illness: 12-month results from a cluster randomized trial of an adaptive implementation strategy. Implement Sci. 2014;9:163.

70. Dreischulte T, Donnan P, Grant A, Hapca A, McCowan C, Guthrie B. Safer prescribing–a trial of education, informatics, and financial incentives. N Engl J Med. 2016;374(11):1053–64.
71. Hargreaves JR, Copas AJ, Beard E, Osrin D, Lewis JJ, Davey C, Thompson JA, Baio G, Fielding KL, Prost A. Five questions to consider before conducting a stepped wedge trial. Trials. 2015;16:350.
72. Brown C, Lilford R. The stepped wedge trial design: a systematic review. BMC Med Res Methodol. 2006;6:54.
73. Davey C, Hargreaves J, Thompson JA, Copas AJ, Beard E, Lewis JJ, Fielding KL. Analysis and reporting of stepped wedge randomised controlled trials: synthesis and critical appraisal of published studies, 2010 to 2014. Trials. 2015;16:358.
74. Snooks H, Bailey-Jones K, Burge-Jones D, Dale J, Davies J, Evans B, Farr A, Fitzsimmons D, Harrison J, Heaven M, et al. Fitzsimmons D, Harrison J, Heaven M et al: Health services and delivery research. In: Predictive risk stratification model: a randomised stepped-wedge trial in primary care (PRISMATIC). Southampton: NIHR Journals Library; 2018.
75. Park JJH, Siden E, Zoratti MJ, Dron L, Harari O, Singer J, Lester RT, Thorlund K, Mills EJ. Systematic review of basket trials, umbrella trials, and platform trials: a landscape analysis of master protocols. Trials. 2019;20(1):572.
76. Edwards H, Redaniel MT, Opmeer B, Peters T, Margelyte R, Sillero Rejon C, Hollingworth W, Craggs P, Hill E, Redwood S, et al. Evaluating an enhanced quality improvement intervention in maternity units: PReCePT trial protocol. BMJ Open Quality. 2021;10(2):e001204.
77. Halpern D, Mason D. Radical incrementalism. Evaluation. 2015;21:143–9.
78. Haynes L, Service O, Goldacre B, Torgerson D. Test, learn, adapt: developing public policy with randomised controlled trials. London: Cabinet Office Behavioural Insights Team; 2012.
79. Ivers NM, Grimshaw JM. Reducing research waste with implementation laboratories. Lancet. 2016;388(10044):547–8.
80. Grimshaw JM, Ivers N, Linklater S, Foy R, Francis JJ, Gude WT. Reinvigorating stagnant science: implementation laboratories and a meta-laboratory to efficiently advance the science of audit and feedback. BMJ Qual Saf. 2019;28(5):416–23.
81. Stanworth SJ, Walwyn R, Grant-Casey J, Hartley S, Moreau L, Lorencatto F, Francis J, Gould N, Swart N, Rowley M, et al. Effectiveness of enhanced performance feedback on appropriate use of blood transfusions: a comparison of 2 cluster randomized trials. JAMA Network Open. 2022;5(2):e220364–e220364.
82. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. Lancet. 2009;374:86–9.
83. Ivers NM, Grimshaw JM, Jamtvedt G, Flottorp S, O'Brien MA, French SD, Young J, Odgaard-Jensen J. Growing literature, stagnant science? Systematic review, meta-regression and cumulative analysis of audit and feedback interventions in health care. J Gen Intern Med. 2014;29(11):1534–41.
84. Karanatsios B, Prang K-H, Verbunt E, Yeung JM, Kelaher M, Gibbs P. Defining key design elements of registry-based randomised controlled trials: a scoping review. Trials. 2020;21(1):552.
85. Rushforth B, Stokes T, Andrews E, Willis TA, McEachan R, Faulkner S, Foy R. Developing 'high impact' guideline-based quality indicators for UK primary care: a multi-stage consensus process. BMC Fam Pract. 2015;16(1):156.
86. Foy R, Eccles MP, Hrisos S, Hawthorne G, Steen N, Gibb I, Croal B, Grimshaw J. A cluster randomised trial of educational messages to improve the primary care of diabetes. Implement Sci. 2011;6:129.
87. McLintock K, Russell AM, Alderson SL, West R, House A, Westerman K, Foy R. The effects of financial incentives for case finding for depression in patients with diabetes and coronary heart disease: interrupted time series analysis. BMJ Open. 2014;4(8):e005178.
88. Ivers N: Supporting audit and feedback to encourage vaccine uptake (NCT05099497). ClinicalTrialsgov 2021.
89. Yu LM, Bafadhel M, Dorward J, Hayward G, Saville BR, Gbinigie O, Van Hecke O, Ogburn E, Evans PH, Thomas NPB, et al. Inhaled budesonide for COVID-19 in people at high risk of complications in the community in the UK (PRINCIPLE): a randomised, controlled, open-label, adaptive platform trial. Lancet. 2021;398(10303):843–55.
90. Foy R, Skrypak M, Alderson S, Ivers NM, McInerney B, Stoddart J, Ingham J, Keenan D. Revitalising audit and feedback to improve patient care. BMJ. 2020;368:m213.
91. Ruppertsberg AI, Ward V, Ridout A, Foy R. The development and application of audit criteria for assessing knowledge exchange plans in health research grant applications. Implement Sci. 2014;9:93.
92. Wilkinson E. RECOVERY trial: the UK covid-19 study resetting expectations for clinical trials. BMJ. 2020;369:m1626.
93. James M, Ford GA, Robinson T, Al-Shahi Salman R. Rapid mobilisation of research in response to covid-19: a paradigm for the future. BMJ. 2020;369:m2155.
94. Wieseler B, Neyt M, Kaiser T, Hulstaert F, Windeler J. Replacing RCTs with real world data for regulatory decision making: a self-fulfilling prophecy? BMJ. 2023;380:e073100.

Publisher's Note