



Evaluating complex interventions: one size does not fit all

Although planning of new health policy could be improved to enable more robust evaluation, **Mhairi Mackenzie and colleagues** argue that randomised controlled trials are not always suitable or practical

In 2000, Medical Research Council guidance recommended that evaluators adopt a sequential approach to testing complex interventions in health care.¹ The approach would lead to a well theorised and replicable intervention that could be assessed using a randomised controlled trial. The model, largely reflecting that adopted in clinical drug trials, was criticised on several fronts, including a failure to appreciate the complexity of policy related programmes and contextual variation. Although the updated framework in 2008 addressed many of these criticisms, it still argued that evaluators should strive to use the model of randomised controlled trials.²

The recent health select committee report on health inequalities³ also criticised the missed opportunities to conduct controlled studies of recent policy interventions and called for policy makers to develop interventions that could be better evaluated. These would be more clearly defined, reasonably stable over time, and have specified levels of consistency in implementation between different contexts. Ideally, these features would provide the opportunity for randomised testing.

However, Pawson and Tilley have argued strongly that treating complex programmes as single interventions is misguided and that the randomised controlled design is not appropriate for answering pertinent questions about what works for whom in what circumstances.⁴ Hawe and Shiell, although advocating the judicious use of controlled designs, have argued that the MRC guidance does not acknowledge the unpredictability of organisational systems into which interventions are introduced. They suggest that, rather than viewing interventions as discrete packages, they should be viewed as “events in systems.”⁵ We use the example of Keep Well (the Scottish government’s major investment in cardiovascular anticipatory care) to show the problems of implementing the MRC recommendations for national policy initiatives.

Keep Well programme

Launched in 2006, Keep Well aims to tackle inequalities in cardiovascular morbidity and mor-

tality in Scotland.⁶ Delivered through primary care, the programme has adopted an anticipatory care approach, modelled on proactive case finding and screening for clinical risk factors, as pioneered in the Netherlands and Wales (box 1).^{7,8}

Keep Well exemplifies complexity as described by the MRC guidance: it has multiple outcomes, many stakeholders, and long chains of hypothesised activity between inputs and outcomes. Furthermore, it operates in a complex system, resulting in the need for adaptation to change in local environments and non-predictability in its behaviour.¹⁰⁻¹²

Hawe and colleagues provide a compelling argument that controlled trial designs are possible in complex interventions when the form (that is, the means of intervening) varies according to local circumstances but the function (or theorised mechanism) of the intervention remains unchanged over place and time.¹¹ Keep Well, however, varies in both form and function across and within pilot sites and over time, reflecting the realities of implementing national policy across differing sites.¹³

For example, differences in stakeholders’ definitions of key theoretical concepts have been identified. Some policy makers argued that anticipatory care is synonymous with health promotion and can be undertaken at a population level by health improvement practitioners; others, however, argue that its defining feature is the empowering and therapeutic relationship generated between patient and general practitioner. This has led to debates about whether the intervention should have been managed at a community level or embedded within general practice consultations. Furthermore, although pilot areas were selected systematically on the basis of their socioeconomic profile, the pilots selected practices in different ways. As a result they do not have equivalent population sizes or concentrations of deprivation within participating practices.

Turning to the intervention itself, the different pilot schemes and practices have not used uniform definitions of perceived need to prioritise their target populations. Some have first tried to engage the least hard to reach individuals, while

others have approached groups thought less likely to access services. Furthermore, all areas have used a growing array of approaches (including sending letters to patients with fixed appointment times, using non-NHS agencies to phone patients to arrange suitable times in convenient venues, and the provision of outreach services) to engage their target populations, but these have been introduced non-systematically within and between pilots and address different rationales for why individuals may be hard to reach. These variations multiply once the programme tries to engage and retain individuals in interventions.

The monitoring data also vary between pilots and, because of non-standard information technology systems and different governance arrangements, there is inconsistency in the timing and type of data available to the national evaluation. With each successive wave of pilots the programme has changed. Keep Well has broadened out from the original pilots across additional geographical areas and has been applied in other settings with different hard to reach populations, greatly reducing the opportunity to identify comparison populations. Crucially, evaluators have little or no control over these types of policy refinements, which are typical of public health interventions.¹⁴

What’s the problem?

Many academics have argued that policy makers have a moral duty to develop policy in a way that allows robust assessments of relative merit and cost¹⁵; the MRC guidance and the health select committee report exemplify this position. However, we argue that a distinction should be drawn between elements of policy interventions that can be shaped for experimentation and those that are inherently problematic.

Several elements of Keep Well could have been better developed and standardised in order to make it easier to evaluate. We give three examples here.

- Data from the national evaluation suggest that some components of the intervention were undertheorised—for example, the programme explicitly targeted hard to reach groups,

but without clarifying who the groups were or the mechanisms by which they became underserved by health services. This is important because it leads to different and potentially conflicting approaches to targeting and reaching the Keep Well population and suggests different functions at play

- Greater efforts could have been made to standardise the function of the reach and engagement strategies (box 1) so that, at least within individual pilots, it would have been possible to describe how different approaches affected the target population
- Larger and earlier investments in robust monitoring systems would have greatly increased the capacity to track causal pathways in real time, both before and during the implementation of Keep Well, as recommended for the evaluation of patient safety interventions.¹⁶

Nonetheless, even with these improvements (which would have required considerable policy and practice commitment and a considerable time delay in establishing the programme), Keep Well would not have reached sufficient standardisation to carry out a controlled trial. Indeed, there are several reasons why advocacy of such an approach is not always appropriate.

Firstly, standardisation does not take into account the well established gulfs between policy as a statement of intent and actual practice. Professionals have been shown to practise in ways that might significantly impinge on the way interventions are expected to work. Adoption of new ways of working may be total, partial, or skewed, particularly in professions with a traditionally strong power base within the NHS, such as general practitioners.¹⁷

Secondly, the MRC framework assumes that interventions that follow the guidance will reach a point of stability. However, complex organisa-

tional systems are characterised by flux, contextual variation, and adaptive (or even maladaptive) learning rather than stability. In Keep Well learning between and during the pilots led to practices changing their approaches over time. Encouraged to operate as reflexive learning organisations, pilot practices met regularly to share learning about how to encourage attendance at health checks and make iterative changes to practice accordingly. This kind of learning occurs independently of the evaluator, and the tension between stability and learning is found within almost all programmes that are implemented in real life rather than within the bounded (and artificial) constraints of a randomised controlled trial. It results in much larger departures from intervention protocols than seen for interventions set in a less complex organisational system.

Thirdly, it is impossible to divorce an intervention from its policy context. This raises a particular problem for the MRC recommendation of using



Box 1 | Summary of Keep Well programme

Aim

To increase the rate of health improvement for those living in the most socioeconomically deprived communities of Scotland through early intervention with those at a high risk of coronary heart disease and diabetes

Target population

Communities with high levels of multiple deprivation and, within these, patients aged 45-64 and registered with participating general practices. The first wave (2006-9) targeted communities in five areas, with a total target population of 87 440. Funding was subsequently extended until 2010.

Approach to anticipatory care

- Identify and target those at risk of preventable serious ill health (including people with undetected chronic disease)
- Invite individuals to attend a health check
- Offer evidence based interventions and services (pharmacological, behavioural, and social) within primary or secondary care, or outside the NHS
- Provide monitoring and follow-up.

Roll-out

Three subsequent waves of Keep Well have been rolled out in new areas. Anticipatory care approaches modelled on Keep Well are also being tested in new settings (eg, community pharmacy), new populations (eg, prisons and black and ethnic minority communities), and in rural or remote areas of Scotland.

Evaluation

The evaluation comprised two phases:

- A theories of change⁹ approach was adopted to delineate the rationale for the programme and to track and test change over time. As a before and after approach was not possible, propensity score matching was explored to identify non-participating practices that can be matched with participating practices on key variables—eg, size of practice and level of deprivation in population
- Informed by principles of realistic evaluation,⁴ a series of case studies has been developed to understand why contextual variations and different approaches to reach and engagement trigger different outcomes

stepped wedge designs to overcome the problem of withholding interventions from control populations. The assumption of non-contamination of controls does not hold true for complex programmes that are inextricably part of a particular policy approach. In this case, “control” practices will have been aware of Keep Well and may well have been making anticipatory adjustments.¹⁸ Any attempt to reduce the potential for contamination is, however, antithetical to the instincts of policy makers, making it impossible to separate entirely the effects of interventions from that of other policy drivers with similar mechanisms.¹⁹

Fourthly, the health select committee report implies that the (strained) public purse should not fund interventions which cannot be evaluated by randomised controlled trial. This presupposes that intervening is a straight alternative to not intervening. In fact, diverse and untested interventions can be found in many public services.

Fifthly, the guidance assumes that policy makers use evidence about the effectiveness of interventions to make instrumental decisions about future action. The role of evidence in policy making is, however, more diffuse, with future policy and action informed as much by formal and experiential learning.^{19,20}

The final issue is one of context. Recognising that not all contexts can be standardised, the MRC guidance accepts a “specified degree of adaptation to local settings” within the research protocol. Arguably, that is an overly narrow conceptualisation of the role of context. Persuasive arguments from theory based evaluation suggest that context is dynamic and integral to learning about why components of interventions trigger change in some individuals or organisations but not in others.^{4,5,9} Experimental approaches are not always suited to generating this type of learning.

What's the solution?

Currently there are no evaluation approaches that are fit for all purposes. Over-standardisation of complex interventions is in danger of delivering precise but invalid effect sizes, while approaches that aim to understand complexity can rarely give definitive answers about whether a complex intervention is effective at the population level. Despite considerable progress in the development of mixed methods and theory based approaches, the lobby for controlled trial designs remains powerful. As a result of the complexity and constantly evolving nature of public policy programmes, such designs are not always possible or appropriate.

Evaluators, policy makers, and research commis-

Box 2 | Examples of questions examined in evaluation of Keep Well¹³

- Was Keep Well implemented as planned and did it meet its expected goals?
- How do contextual variations relate to implementation and impact?
- To what extent did Keep Well reach and engage the most socioeconomically deprived populations and were these the most at risk clinically?
- To what extent did Keep Well become normal standard practice?
- How are outreach approaches conceptualised and implemented?
- How do differences in approach relate to contextual differences and is impact related to approach?

sioners need to encourage greater conceptual clarity at the heart of their complex interventions, more robust data collection systems, and more theoretically driven questions that seek to understand and work with context in more meaningful ways. These recommendations are consistent with the MRC guidance, and, to the extent that the MRC guidance pushes for more rigorous evaluative thinking in general, it is a welcome starting point.

However, most policy interventions are of such complexity that it is counterproductive to view randomised controlled trials as the best method of assessment. Nor is it realistic to argue that policy makers should develop only interventions amenable to such approaches. Finally, there are very many evaluation questions that are worthy of answering beyond that of whether an intervention works (box 2). These may seem less ambitious questions than those that can be answered through a controlled trial, but they are, nonetheless, highly pertinent to health improvement.

Mhairi Mackenzie senior lecturer in public policy, Department of Urban Studies, University of Glasgow, Glasgow G12 8RS

Catherine O'Donnell professor of primary care research and development, General Practice and Primary Care, Division of Community based Sciences, University of Glasgow

Emma Halliday public health adviser, Policy Evaluation and Appraisal, NHS Health Scotland, Edinburgh EH12 5HE

Sanjeev Sridharan associate professor, health policy, management and evaluation, Keenan Research Centre, St Michael's Hospital, University of Toronto, Toronto, Ontario, Canada

Stephen Platt professor of health policy research, Centre for Population Health Sciences, University of Edinburgh, Edinburgh, EH8 9AG

Correspondence to: M Mackenzie m.mackenzie@lbss.gla.ac.uk
Accepted: 29 December 2009

Contributors and sources: The authors all have extensive experience in conducting or commissioning evaluations of complex health policy interventions. All except EH are part of the National Evaluation of Keep Well. EH is the lead commissioner of

the evaluation of Keep Well and works within a policy evaluation and appraisal team with responsibility for bridging policy, practice, and evaluation. MM took lead responsibility for drafting the article and will act as guarantor; all other authors had a significant role in its critical revision.

We thank Trisha Greenhalgh, Sally Wyke, and Richard Lilford for their encouraging and constructive comments.

Funding: The research on which this paper is based is funded by NHS Health Scotland and the health improvement strategy division in the chief medical officer and public health directorate of the Scottish Government. Opinions expressed in the paper are those of the authors and do not necessarily reflect the views of NHS Health Scotland.

Competing interests: None declared.

Provenance and peer review: Not commissioned; externally peer reviewed.

- 1 Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. *BMJ* 2000;321:694-6.
- 2 Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. *Developing and evaluating complex interventions: new guidance*. London: Medical Research Council, 2008.
- 3 House of Commons Health Committee. Health inequalities third report of session 2008-09. Stationery Office, 2009.
- 4 Pawson R, Tilley N. *Realistic evaluation*. Sage Publications, 1997.
- 5 Hawe P, Shiell A, Riley T. Theorising interventions as events in systems. *Am J Community Psychol* 2009;43:267-76.
- 6 Scottish Government. Delivering for health. Stationery Office, 2005. www.scotland.gov.uk/Publications/2005/11/02102635/26356.
- 7 Crebolder HFJM, van der Horst FG. Anticipatory care and the role of Dutch general practice in health promotion. *Patient Educ Counselling* 1996;28:51-5.
- 8 Hart JT, Thomas C, Gibbons B, Edwards C, Hart M, Jones J, et al. Twenty five years of case finding and audit in a socially deprived community. *BMJ* 1991;302:1509-13.
- 9 Connell JP, Kubisch AC, Schorr LB, Weiss CH. *New approaches to evaluating community initiatives. Volume 1: concepts, methods and contexts*. Washington, DC: Aspen Institute, 1995.
- 10 Shiell A, Hawe P, Gold L. Complex interventions or complex systems? Implications for health economic evaluation. *BMJ* 2008;336:1281-3.
- 11 Hawe P, Shiell A, Riley T. Complex interventions: how “out of control” can a randomised controlled trial be? *BMJ* 2004;328:1561-3.
- 12 Rogers PL. Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation* 2008;14:29-48.
- 13 O'Donnell K, Mackenzie M, Platt S, Sridharan S, Reid M, Turner F, et al. National evaluation of Keep Well: interim report national level. University of Glasgow, 2008.
- 14 Sowden SL, Raine R. Running along parallel lines: how political reality impedes the evaluation of public health interventions—a case study of exercise referral schemes in England. *J Epidemiol Community Health* 2008;62:835-41.
- 15 Oakley A. Experimentation and social interventions: a forgotten but important history. *BMJ* 1998;317:1239-42.
- 16 Brown C, Lilford R. Evaluating service delivery interventions to enhance patient safety. *BMJ* 2008;337:a2764.
- 17 May C, Finch T, Mair F, Ballini L, Dowrick C, Eccles M, et al. Understanding the implementation of complex interventions in health care: the normalization process model *BMC Health Serv Res* 2007;7:148.
- 18 Campbell S, Reeves D, Kontopantelis E, Middleton E, Sibbald B, Roland M. Quality of primary care in England with the introduction of pay for performance. *N Engl J Med* 2007;357:181-90.
- 19 Weiss CH. *Evaluation: methods for studying programs and policies*. 2nd ed. Prentice Hall, 1998.
- 20 Petticrew M, Whitehead M, Macintyre SJ, Graham H, Egan M. Evidence for public health policy on inequalities: 1. The reality according to policymakers. *J Epidemiol Community Health* 2004;58:811-6.

Cite this as: *BMJ* 2010;340:c185