Chapter 1

Introduction

1.1 Evaluating Complex Health Interventions

This book is about understanding the impacts of health interventions. By interventions, we mean actions that are purposefully taken to bring about specific, intended benefits. These actions could include implementing policies, modifying services, launching new social programmes or finding other new ways of working with patients or citizens. We are interested in *complex* health interventions. These are interventions that are more than just giving a pill or doing a particular medical or surgical procedure. They involve multiple components that interact with each other and with the context in which they are delivered.[1] Our own scientific research focuses on public health interventions, which aim to prevent disease or promote population health. Most of the public health challenges that we face today, such as preventing violence, obesity, infectious diseases or mental illness, are complex. They require interventions to address multiple influences on health operating at the level of individuals, communities and whole societies. These interventions might involve communicating health messages, providing people with support, persuading people to change their behaviour or changing the environments in which people live to make healthier decisions easier.

These interventions are complex firstly because they have multiple components that interact with each other.[2] The harder it is to say what the 'active ingredient' of an intervention is, the more likely it is to be a complex intervention.[3] Consider the 'Intervention with Microfinance for AIDS and Gender Equity' intervention. This aimed to reduce HIV infections and gender-based violence among poor women and their family members in rural South Africa. It did so by providing workshops which educated the women about HIV and gender, empowering the women by enabling them to work together to lead campaigns in their communities on issues of importance to them and providing the women with small loans to start small businesses and ease their poverty.[4] The intervention developers believed that all these components would work together so that women had the knowledge, motivation and life circumstances necessary to reduce their own and their family members' risk of HIV and gender-based violence. In other words, they hypothesised that the impact of the overall intervention would be greater than the sum of its parts.[5] But it is not only public health interventions like these that can be complex. Healthcare interventions can also be complex, even when they may not appear to be. The care that patients receive is not usually just limited to a single pill or procedure. It involves various activities which interact with each other. For example, interventions to remind pregnant women to take up a glucose tolerance test for gestational diabetes can include 'cues to action' for providers, different

https://doi.org/10.1017/9781009456616.001 Published online by Cambridge University Press

types of reminders for pregnant women and the provision of resources to facilitate women's ability to use a glucose tolerance test.[6]

Secondly, a clue that interventions could be complex is if they work differently across different contexts, that is with different populations or in different settings.[2 7] How interventions are delivered and the impact they have will depend on local factors, such as whether they are supported by local policies, whether potential implementation partners are ready to deliver them, whether they reach local people and whether they meet the needs of those they are meant to benefit.[8] The local capacity to implement an intervention and the local capacity to benefit from a complex intervention will vary. Consider the example of using youth service interventions as a way to prevent teenage pregnancies. The intervention might involve a youth worker mentoring young people, giving young people additional education on academic and life skills and facilitating group activities to build self-esteem and raise aspirations. Interventions of this sort have been found to reduce teenage pregnancy rates in New York City but not in other parts of the USA.[9 10] Some of us evaluated a government-led pilot programme to implement this sort of intervention in England and found that it actually *increased* the rate of teenage pregnancy.[11] Various factors might explain these differences in impact across different contexts. In England, the youth work was sometimes provided as an alternative rather than a complement to normal schooling so the students referred into it might have felt like their involvement labelled them as failures. The youth work itself sometimes was delivered with low fidelity so it did not match up to what was intended. In New York City, the intervention was delivered to all young people living in areas of poverty. However, in England the intervention was targeted to particular young people whom teachers or social workers judged as at particularly high risk of pregnancy. The intervention might have increased the rate of pregnancy in this English context by bringing together the most at-risk young people, possibly leading to more sex without contraception.[12] A useful way to think about an intervention is that it is a disruption to an existing social system. By social system, we mean the places, environments and community values and practices that shape what we believe and how we behave. The pre-existing features of this system will shape how the intervention is delivered. And different social systems will be 'disrupted' in different ways by the same sort of intervention, with implications for local impacts.[13] This book will provide an approach to evaluation which rigorously assesses the outcomes of interventions and how these vary between contexts.

1.2 Why Evaluation is Important

In this book, we argue that the evaluation of complex health interventions and the use of this evidence to inform policy are critically important but currently are not achieving their full potential. The reason why evaluation is important is because it is usually not obvious what impacts complex interventions have. Interventions might bring about lots of benefits or they might do nothing, waste money or even harm people. Even if they are beneficial in some contexts, interventions may not work everywhere, as the above example of youth work and teenage pregnancies demonstrates. Interventions are always delivered with good intentions, but it is often not obvious to those delivering or receiving interventions what impacts (if any) they have had. The impacts might be too subtle to be noticed. We don't need evaluations to tell us that parachutes work but most interventions do not generate such dramatic outcomes as do parachutes. It can also be difficult to distinguish intervention impacts from other changes happening at the same time.

https://doi.org/10.1017/9781009456616.001 Published online by Cambridge University Press

One problem is 'regression to the mean'. This occurs, for example, when clients or places receive an intervention for the very reason that, at that moment, they are at heightened risk of some adverse outcome. Their risk naturally fluctuates up and down over time, and it goes down to a lower level at about the same time as they received the intervention without this being a result of the intervention. For example, you might seek an HIV safer sex counselling intervention when you are concerned about your current risk of HIV. You might have gone on holiday and had more sex than usual or used protection less than you normally would. Your level of risk will probably dip back down independently of any impact of the counselling.[14]

Another problem is distinguishing the effects of an intervention on a population from the broader trends affecting that population. There might be 'maturational trends' (people changing as they get older) or 'secular trends' (people being affected by long-term historical changes). Because of these trends, it can be hard for those delivering or receiving an intervention to separate the 'signal' (intervention effects) from the 'noise' (other trends or events). This issue can also challenge evaluators, a subject we will turn to in Chapter 2.

A particularly important role for evaluation is to detect harms. No one wants to continue to deliver an intervention that causes harms. But, like intervention benefits, these may not always be obvious. 'First do no harm' is an ethical requirement that ranks higher even than doing good.[15] Although it is easy to imagine how medicines or surgery could inadvertently cause patients harm, it is harder to imagine that interventions such as education, social support or environmental improvements might cause harm. Unfortunately, however, lots of evidence indicates that this can sometimes happen.[12] As well as the example of youth work and teenage pregnancy described in the previous section, another classic example is that of the Cambridge-Somerville social work intervention. This involved providing a broad set of social work interventions, such as counselling and free places on summer camps, for at-risk boys in New England, USA, in the 1930s. Those who received the intervention were found to experience higher rates of criminal activity, alcoholism and mental illness later in life.[12] Because interventions are interruptions to complex social systems, it is plausible that unintended effects can occur, some of which might be harmful.[16] The assessment of harms has been a neglected topic in evaluations of public health interventions,[17] other than for a few topics such as suicide prevention and illicit drug interventions.[18 19] But interest in the potentially harmful effects of public health intervention has increased recently and researchers have tried to define different categories of harm.[12 17] One way to do this is to distinguish between 'paradoxical effects' (the intervention making worse the very thing it is trying to make better) and 'harmful externalities' (the interventions bringing about harms in completely different areas).[20]

Interventions often aim to reduce health inequalities. These are avoidable, unnecessary and unfair differences between groups in health status and outcomes. These may arise as a result of the unequal distribution of resources or as a result of discrimination or other unequal access to rights. Minoritised and racialised groups experience worse health outcomes across a range of conditions. Women experience disproportionate impacts from intimate partner violence. People experiencing poverty are less able to access health services. Interventions may often aim to reduce health inequalities, but some will actually increase health inequalities, benefiting the health of the advantaged more than the disadvantaged even if the health of no individuals is directly harmed by an intervention. When interventions disproportionately benefit people who already have better health, we call this an 'equity harm'.[17] Conversely, interventions that decrease gaps between groups in health

https://doi.org/10.1017/9781009456616.001 Published online by Cambridge University Press

status can be said to create 'equity benefits'. Certain types of interventions, such as mass media interventions, are known for being more likely to create equity harms because only those already most able to take up mass media messages do so.[21] Our book will identify approaches that ensure that evaluation can rigorously assess not only whether interventions achieve their intended effects but also whether they generate any harmful effects.

But evaluation is expensive, complicated and time-consuming. We cannot evaluate every intervention all of the time to make sure that it is benefiting and not harming those it aims to help. We need to decide when to evaluate interventions and when not to bother. If the intervention has dramatic, obvious effects, an evaluation is not needed unless there are concerns of possible harmful externalities. If an intervention is cheap, easy to deliver, acceptable and there is minimal risk of harm, it may also not be worth evaluating.[22 23] It may also not be worth evaluating an intervention if it is delivered as a one-off with no plans to repeat it over time or in different places. But an intervention might be worth evaluating if its expected impacts are subtle; if it is costly, difficult or controversial to deliver; if it has the potential for harmful effects; and if it will be delivered in more than one time or place.

A single evaluation study is unlikely to provide a definitive guide to whether an intervention is a potentially useful approach to use across contexts. The results of a single study may be biased by limitations in the methods used or the biases of those leading the evaluation. A single evaluation undertaken at a single point in time and in a single place is unlikely to provide evidence that will allow us to decide where else and for whom else the intervention should be delivered. So we usually need multiple studies to better understand intervention effects. The results of these individual studies need to be critically appraised and their results summarised in what are known as 'systematic reviews'. In Chapter 2, we describe conventional approaches to evaluation and systematic reviews, and some of the limitations with these conventional approaches.

1.3 The Value of Evidence in Informing Policy

Evaluation and the use of evidence to inform policy have a long history. Authors such as Donald Campbell, Robert Merton and Karl Popper, writing in the mid-twentieth century, argued that we need experiments to inform and assess government policies.[24–26] Popper termed this 'piecemeal social engineering', meaning incremental changes to policies or services which are then evaluated to assess whether they have the intended impacts or whether they have caused unintended harms. At this time, there were only a few examples of large-scale evaluations in areas such as agriculture, education, medicine and social work (including the Cambridge-Somerville study). Most policy decisions were made on the basis of tradition, political ideology or simply the views of those in charge. The last of these is well illustrated by the statement that 'the gentleman [sic] in Whitehall really does know better what is good for people than the people know themselves'[27] (p. 317).

Popper saw evaluation and the basing of policy on evidence of impact as a way to resolve tensions between conservative, liberal and socialist ideologies that were playing out between the eighteenth and twentieth centuries.[25] Conservatives thought that societies should stick with traditional ways since these were tried and tested, representing the collective wisdom of previous generations. Liberals wanted new policies to improve social conditions and promote individuals' welfare and rights. Socialists demanded or anticipated radical changes to how the economy was run to make societies fairer. Popper argued that radical policy change was often grounded in theories about society for which there was no evidence. These

https://doi.org/10.1017/9781009456616.001 Published online by Cambridge University Press

policies could bring about unintended harms (such as tyranny, violence and mass starvation). The speed and scale of these policy changes could leave insufficient time for evaluation or improvement. Popper, as well as Campbell and Merton, recommended that social policies should focus on gradual change and empirical evaluation of their effects. Karl Popper proposed piecemeal social engineering informed by experimentation.[25] Robert Merton argued for the importance of developing scientific theory informed by evidence to guide policy.[26] Donald Campbell coined the term 'the experimenting society' as a way to think of how policy change could progress based on careful trial and error.[28] Popper viewed social democracy as the form of government which could gradually address the inequalities generated by capitalism, ensuring that citizens received education, health and welfare, and were entitled to civil and worker rights.[25] In liberal democracies, the idea that policy should be based on evidence started to become more popular in the 1960s and became really influential from the 1990s. During this period, when centrist and centre-left governments ran many countries, evidence-informed policy came to be associated with a 'technocratic, Third Way' approach, summed up in Tony Blair's phrase 'what matters is what works'.[29]

Some critics argue that the use of evaluation and other research evidence to inform policy and practice is merely part of the apparatus through which the state and experts control service providers and citizens.[30] They argue that using evidence in this way narrows policymaking to a series of expert-led technocratic assessments squeezing out democratic consideration of values and priorities. Evidence, it is argued, can be a way to obscure the political way in which the powerful decide what counts as a 'problem' or a plausible 'solution'.[31] Quantitative evidence is regarded by some critics as particularly problematic because, it is argued, it tends to prioritise precision (in estimating what factors cause or what interventions affect health outcomes) over depth (in terms of analysing the broader social structures which bring these problems about).[32 33] It is argued that the use of insufficiently 'upstream' analyses then informs the use of insufficiently upstream interventions so that the deeper causes of health inequalities remain unexamined and unchallenged.[32 33] We disagree with such analyses; the use of evidence from evaluations and other research need not bring about undemocratic and de-politicised decision-making. There is no inevitable trade-off between statistical precision and depth of analysis.[32] Use of quantitative evidence need not preclude the assessment of how deeper social forces contribute to health inequalities or the impacts of interventions addressing these forces.[34 35] In this book, we offer recommendations for how evaluation can contribute to evidence-based policymaking in more useful ways than has been achieved to date.

1.4 The Strengths and Current Limitations of Randomised Controlled Trials and Systematic Reviews

We strongly support the use of randomised controlled trials, (or trials for short), where possible, to assess the impacts of complex health interventions. We also strongly support the use of systematic reviews to collate evidence from multiple studies and using this to inform policy decisions. We believe that trials and systematic reviews offer the most scientifically rigorous means of determining the impacts of interventions. Trials produce the least biased statistical estimate of how much better are the outcomes of people who are allocated to receive an intervention compared to those who are allocated not to receive this. Systematic reviews collate evidence from various studies to provide the most comprehensive answer to

https://doi.org/10.1017/9781009456616.001 Published online by Cambridge University Press

the question of whether or not an intervention 'works'. In Chapter 2, we explain why this is so.

However, while we believe that randomised trials and systematic reviews, when done well, are very scientifically rigorous as methods, we argue in this book that they are often not scientific enough in their overall orientation, that is what questions they ask and what evidence is used for. They generally focus on questions of *if* and *how much* interventions work, and generally do not focus enough on understanding *how* interventions work and *for whom* or *where* they work best. This is a critical gap in the evaluation of complex health interventions because, by definition, these interventions work via complex mechanisms, which also interact with local context, to generate different impacts in different places or populations. Because of the failure to consider context and mechanism in meaningful ways, evaluators cannot provide policymakers or practitioners with the evidence that they need to decide if the intervention in question may be beneficial beyond the context of the original evaluation. In Chapter 3, we argue that this seriously limits the usefulness of evaluation evidence in informing policy.

Some people argue that the reason trials are not very useful is that they try to apply methods from the natural sciences to understand how the social world works. Complex interventions involve interactions between people, with people deciding how to change their actions based on their understanding and experience of an intervention. Critics argue that trials (and science more generally) are just not appropriate to understanding this messy and nuanced social world.[36] We disagree with this position. Instead, we argue in Chapter 3 that the evaluation of complex interventions actually needs to become more, not less, scientific. Currently, evaluation, including trials, is generally limited to being a sophisticated form of intervention monitoring. Evaluations *describe* the impacts of an intervention statistically and use this as a basis for 'accrediting' interventions as effective or not (in Tony Blair's terms 'what works'[29]). Instead, we argue, trials need to contribute to testing and refining scientific theories about how and for whom interventions work.

In Chapter 3, we draw on ideas from 'realist' evaluation methods to develop a method by which trials can become more scientific. We call our method 'realist trials'. We describe this method in detail in chapters 4, 5 and 6. In chapters 7 and 8, we consider how the method can also be applied to improving systematic reviews. We call this approach 'realist systematic reviews'. In Chapter 9, we consider how our methods can help make evidence more useful in informing policy decisions. In Chapter 10, we consider how our methods can be used to test and refine scientific theories, which might then in turn be used to inform interventions and policy in the longer term. Our ideas are controversial and some researchers disagree with our approach.[37] [38] But we hope to present arguments and evidence to show that ours is the right approach.

https://doi.org/10.1017/9781009456616.001 Published online by Cambridge University Press