

Evaluation

<http://evi.sagepub.com/>

Evaluating complex and unfolding interventions in real time

Tom Ling

Evaluation 2012 18: 79

DOI: 10.1177/1356389011429629

The online version of this article can be found at:

<http://evi.sagepub.com/content/18/1/79>

Published by:



<http://www.sagepublications.com>

On behalf of:

THE
TAVISTOCK
INSTITUTE®

The Tavistock Institute

Additional services and information for *Evaluation* can be found at:

Email Alerts: <http://evi.sagepub.com/cgi/alerts>

Subscriptions: <http://evi.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations: <http://evi.sagepub.com/content/18/1/79.refs.html>

>> [Version of Record](#) - Jan 22, 2012

[What is This?](#)



Evaluating complex and unfolding interventions in real time

Evaluation
18(1) 79–91

© The Author(s) 2012
Reprints and permission: sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1356389011429629
evi.sagepub.com



Tom Ling
RAND Europe, UK

Abstract

This article discusses an approach to managing the evaluation of complex interventions. Complex interventions pose significant challenges to the role and conduct of evaluations. In particular, they combine with reflexive learning and change to produce significant uncertainties making it hard to describe in advance what the intervention will do or what the outcomes might be. These uncertainties vary in nature and addressing these uncertainties leads to different evaluation approaches. That such evaluations often take place in ‘real time’ and have a strong formative dimension adds to the challenge. This article proposes a way of approaching this problem by incorporating the concepts of uncertainty and complexity into a Theory of Change-based approach.

Keywords

complexity, evaluation, real-time evaluation, uncertainty

Introduction

The issue to be explored in this article will be familiar to most of the readers of *Evaluation* as the journal has a strong record of considering this issue (see Pederson and Reiper, 2008; Rogers, 2008; Sampson, 2007; Sanderson, 2000); the issue is how to evaluate a complex intervention as it emerges and unfolds, often in unanticipated ways, in an evaluation progressing in ‘real time’ contemporaneously with the intervention itself. The urge to think about this arises from the author’s involvement in a number of such evaluations (in particular, evaluations of clinicians’ engagement in quality improvement, of the Department of Health’s integrated care pilots, and of efforts to bring together researchers and clinicians to provide improved uptake of more patient-focused research; see Ling et al., 2010a, 2010b; Soper et al., 2008). However, the evidence used to explore the topic is from a range of published accounts.

The aim of the article is, first, to attempt to clarify the relevant terms and issues and, second, to discuss the implications for the conduct of evaluations. So first, we need to clarify our terms.

Corresponding author:

Tom Ling, RAND Europe, Westbrook Centre, Milton Rd, Cambridge, CB4 1YG, UK.
Email: tling@rand.org

Complex interventions and reflexivity

To avoid 'complex intervention' being used to describe many different things, we need to define our terms. Famously, Weaver (1948) distinguished between phenomena that were simple (limited, predictable parts and interactions), complicated (multiple but broadly stable parts and interactions), and organized complexity (multiple parts and interactions which interact in unpredictable but non-random and patterned ways). A number of commentaries since that time have drawn upon this three-part distinction (see, helpfully, Rogers, 2008), but it might be helpful to briefly rehearse these here.

Simple interventions rely upon a single (a coherent set of) known mechanism with a single (a coherent set of) output whose benefits are understood to lead to measurable and widely anticipated outcomes. In healthcare, these might be biological (a new drug), psychosocial (a new behavioural therapy) or service delivery-based interventions. Complicated interventions involve a number of interrelated parts, all of which are required to function in a predictable way if the whole intervention is to succeed. The processes are broadly predictable and outputs arrive at outcomes in well-understood ways. A television is complicated but has predictable and stable outcomes (it does not transform itself over time into a toaster). Complex interventions are characterized by feedback loops, adaptation and learning by both those delivering and those receiving the intervention, they have a portfolio of activities and desired outcomes which may be re-prioritized or changed, and they are both sensitive to starting conditions and outcomes tend to change, possibly significantly, over time. Complex interventions, therefore, have multiple components which may act independently and interdependently (Campbell et al., 2007). This is characteristic of many of the things the evaluation profession is asked to evaluate.

It is often noted that in health care, the reductionist approach, of the Randomized Controlled Trial (RCT), is held to be the ideal deductive method (Martin and Sturmberg, 2009). In an important and influential publication from the Medical Research Council (MRC), it is argued that we can apply the same approach to complex interventions if sufficient care is taken (Medical Research Council, 2008). The framework proposes a series of steps leading to a RCT. However, what they focus on is, in Weaver's (1948) terminology, complicated interventions meaning that the various parts can be thought of as interacting in predictable ways and these interactions can be analysed leading to (universally) generalizable conclusions. Furthermore, by focusing on the intervention, the framework misses the point that interventions interact with complex systems in ways that cannot be predicted. The evaluation challenge lies in understanding this interaction.

Achieving this understanding has often proved difficult. However, this is not because the MRC approach is technically challenging (which they recognize is the case), but because it is inherently unable to deal with complexity. This is a problem both for those implementing complex interventions and those evaluating them. In particular, interventions which pay insufficient attention to complexity seem to start with optimism and end with modest or unmeasurable outcomes (Ling et al., 2010b). Evaluators need to be able to understand why '... when trials with organisations and whole communities do go ahead, the story is consistently becoming one of expensive failure – that is, weak or non-significant findings at huge cost' (Hawe et al., 2004), and 'the challenges associated with changing clinician behaviour to improve care, with most interventions showing only limited or modest success' (Litaker et al., 2006), and we can also note the concerns of the Institute of Medicine about the unintended consequences of health programmes and policies (Institute of Medicine, 2001). Two concepts are fundamental to explaining this experience. First, as mentioned, is complexity. The second is reflexivity.

Practitioners delivering complex interventions learn and adapt, as do groups and individuals in the target population. They reflexively seek to make sense of the systems in which they act and where possible to change how they work. They adapt behaviour based on a changing understanding of the consequences of their actions. Of course the extent to which learning and adaptation overcomes routine and inertia is an empirical question and inevitably they (like evaluators) only have an incomplete understanding of these systems and their actions based on this limited understanding may be unpredictable. The behaviour of social actors is therefore shaped by how they reflect upon these systems and adapt their behaviour in the light of their experiences of them. However, because these systems evolve and interact in often unpredictable ways, this reflexivity fails to deliver either certainty or control. Consequently, as Giddens (1990) argued, the early modern expectation that accumulating knowledge will provide humans with greater control has been undermined. The dilemma is evident; policy makers' engagement with the concepts of joined-up working and complexity is evident in policy arenas such as regenerating communities, improving parenting, and supporting learning but at the same time it has heightened our understanding of the inadequate evidence base to guide such decisions (Sanderson, 2009). This article argues that, while we need to challenge the expectation that evaluations of the complex will lead to more precise predictions and greater control, we should not abandon the belief that appropriately structured evaluations can contribute positively to reflexivity while simultaneously fulfilling the evaluators' mission to strengthen both learning and accountability. To do so we will need to trade our search for universal generalizability in favour of more modest, more contingent, claims. In evaluating complex interventions we should settle for constantly improving understanding and practice by focusing on reducing key uncertainties.

Challenging the certainties offered by reductionist approaches to evaluation

More conventional approaches aim to arrive at simple or complicated (in the sense outlined above) truths through the building block of individual pieces of (irrefutable) evidence. They claim to allow us to understand the whole by seeing it as a combination of all of its parts. The evaluation method is therefore to build up detailed pieces of evidence into an accurate account of the costs (or efforts) and the consequences. More widely we might add up all the inputs, describe the processes, list the outputs and (possibly) weighted outcomes and put this together to form judgements about and draw evaluative conclusions. Indeed this might be an appropriate framework for simple and complicated interventions. The framework is appropriate where we can make highly plausible assumptions that we know enough about both the intervention and the context. In more complex and reflexive cases, however, the underlying limitations of this approach (such as that informing the MRC framework) become apparent.

Complexity thinking, on the other hand, argues that we need to start with an understanding of the systems within which the parts operate (Bettis and Prahalad, 1995) and it brings with it quite different causes of uncertainty. The whole is seen to be logically and causally prior to the parts. If we take an example such as evaluating clinicians' engagement with quality improvement, it is obvious that what matters is not simply aggregating the factors known to help (leadership, project management, incentives, fit with professional identity). In an evaluation it is clear that it is not simply the presence of each of these (and the more the better). Rather, it is how these parts are combined and balanced (often with neither too much nor too little) and how they are shaped to address local circumstances or resonate with national agendas. In other words how they form a

system of improvement and how this system interacts with other systems in and around healthcare services. From an evaluators' point of view, 'What matters is making sense of what is relevant, i.e. how a particular intervention works in the dynamics of particular settings and contexts' (Martin and Sturmberg, 2009). And from a complexity-influenced approach, the dynamics of particular settings and contexts are driven by how systems and subsystems interact.

Howe and colleagues also argue that the MRC framework is really describing 'complicated' interventions. This is about much more than semantics since it changes how we think about building conclusions and how generalizable these conclusions can be. As they say in relation to economic evaluation:

Thus, complex interventions of the sort discussed by the MRC [i.e. complicated interventions] are more difficult to evaluate, but there is nothing substantively different about their economic evaluation. No new economic methods are required, and the problems can all be solved with time, effort, and resources.

In contrast, evaluating the economic efficiency of interventions directed at changing the properties of complex systems presents big challenges. Complex systems have several defining characteristics including the tendency to be self organizing, be sensitive to initial conditions, and make non-linear phase transitions (to jump quickly from one position to another very different position); the existence of emergent properties; and the importance of interaction effects and feedback. These characteristics affect what measures of effectiveness should be included in the economic evaluation and how the consequences of the intervention are valued. (Shiell et al., 2008)

The issue of conceptualizing complex interventions is made more difficult still by the fact that we rarely find an intervention that can adequately be described as a single system. More often there are systems nested within systems. 'Complex service interventions, therefore, can be conceptualized as dynamic complex systems thrust amidst complex systems, relentlessly subject to negotiation, resistance, adaptation, leak and borrow, bloom and fade, and so on' (Pawson et al., 2005). Teacher professional learning, for example, can be analysed as involving systems within systems (e.g. Stoller et al., 2006). Quality improvement in healthcare, for example, takes place within systems and subsystems which are mutually interdependent (Ling et al., 2010b) and these might helpfully be thought of as operating at individual, organization and whole-system levels (or micro, meso and macro). When we talk about an intervention being context-dependent, or context-rich, we are describing how the processes and outcomes in each case are shaped by the particular ways in which these systems and subsystems uniquely interact. This raises particular challenges for any evaluations aiming for certainty about attribution and hoping to provide decision-makers with clear recommendations about the generalizability of the approach. It is not unreasonable for policy makers to ask evaluators questions such as 'is integrated care a good thing' but they should not be too disappointed when the answer begins with the words 'it depends'.

For example, in a discussion of evaluation completed for the UK Stabilisation Unit, Van Stolk and colleagues (2011) noted that intervening to provide stability in places like Afghanistan and Iraq involves moving individuals and groups away from using violent means to pursue power and resolve disputes. Much of the activity aims to change the systems and subsystems by, for example, reducing corruption, changing incentives, or changing expectations and perceptions. This might be done in very many different ways depending upon local circumstances. As one participant observed: stabilization is 'a space rather than a plan. First of all we did not have a plan; we had a broad spectrum of possible courses of action. Each situation is different – either requiring rapid intervention or slower consensus building'. This poses very particular problems

for how to conduct and use monitoring and evaluation (M&E) and the authors summarize these problems as:

- The particular way in which stabilisation interventions tend to unfold, with a wide range of often concurrent activities that have different underlying logics;
- Related to the first, the different time horizons and pressures for measuring progress that apply to the actors in a given stabilisation intervention;
- The limited capacities (e.g. organisational culture and technical capacities) of actors involved in stabilisation to undertake M&E activities, owing to time pressure and lack of training in M&E;
- The complexity of the environment in which stabilisation takes place, where what you are measuring is often intangible, which in turn has an impact on M&E processes such as data collection and the interpretation of data. (Van Stolk et al., 2011)

These bullet points highlight some significant challenges facing real-time evaluations.¹ They develop findings established by Hallam (1998). Most importantly the ‘intangible’ changes taking place are often at the level of the interactions among systems. For example, they concern how diplomatic, military and humanitarian activities interact to reinforce or undercut trust, confidence, and attitudes towards violence. The evidence showing that one set of activities is contributing towards stabilization in one context, may tell us little about whether the same activities would do so in a different context.

This is a real challenge for thinking about the evidence base for an evaluation. Anderson makes a related point in his critique of the MRC position:

As a health economist, I find the recommendations on assessing cost effectiveness disappointingly brief. **Crucially, they do not indicate how the different dimensions of complexity of the intervention challenge existing methods for conducting an economic evaluation.** Also, by repeating the conventional view that “the main purpose of an economic evaluation is estimation rather than hypothesis testing,” the guidance may unwittingly encourage the status quo. Most economic evaluations are still primarily quantitative evaluations of “black box” interventions - that is, with little or no explicit interest in how and why they generate different effects or place different demands on the use of resources – so evidence for explaining differential cost effectiveness is often speculative rather than empirical. This is perhaps unsurprising . . . few attempts have been made to bridge the gap between methods of economic evaluation and the broader methodological debates about the definition and evaluation of complex interventions. (Anderson, 2008, author’s emphasis)

Anderson is concerned particularly with the question of economic evaluation but the point can be more generally made. The question is what this implies for evaluation.

Implications for evaluators

First, evaluations of complex interventions must provide more information about the complex systems which shape the context. We should immediately recognize that the view that context is important is not new. Realist evaluation has long been concerned with bringing in context, but it is by no means alone in this insistence. However, this approach to contextualization could lead to the conclusion that every context is different and unique and so we cannot use the lessons from one evaluation to inform decisions elsewhere (Goldstone, 2006). To address this challenge, we can use

complexity thinking to go beyond simply arguing that each context is different by showing how particular systems function and how systems interact. If this were successful it would provide a way of contextualizing and then allowing 'mid-range generalization'. This could deliver a sufficiently thick description of the workings of systems and subsystems to support reflexive learning within the intervention and more informed decision making elsewhere. It establishes a mid-ground between the uniqueness of everything and universal generalizability.

For example, the construction of professional and service-user identities and behaviours are subject to system-like processes where behaviours are reinforced, or attitudes are adapted (perhaps to reduce cognitive dissonance), which work at the small-group or individual level but are also part of much wider systems of professional training and revalidation, the construction of service users as consumers and so forth. Similarly, efforts to reduce high death rates in people with cardiovascular disease are a part of (rather than an intervention into) the complex systems of: disease (atherosclerosis, risk factors, comorbidity); patient beliefs and behaviours; practitioner behaviour and accessibility; health service organization; public policy; and socioeconomic circumstances (Campbell et al., 2007). Any intervention will always be locked into these systems and we need to understand how this facilitates or undercuts different options. In the case of evaluating stabilization activities in Afghanistan, for example, it allows us to draw upon an understanding of how military, humanitarian and diplomatic efforts might mutually reinforce each other or, conversely, how they might systematically undermine each other.

But is this a radically different approach to evaluation? As Kernick and colleagues note:

a more realistic perspective is to see complexity theory complementing existing approaches but alerting us to the importance of matching the research approach to the context and complexity of the environment to which it is applied. An important first step is to recognize the limitations of the dominant research discourses - the assumption that order needs to be created by external forces and that the certainty of structures seen in hindsight offer a central understanding for the emergent order that frames living forward. To be more receptive to approaches that view research not as an endeavour to configure the health service against detailed criteria, but to establish the context where the system will self-organize within a framework of broader policy objectives. (Kernick, 2006)

The first claim is that complexity thinking can complement other approaches to framing evaluations. This suggests that complexity can somehow be grafted onto more conventional approaches. However, at an epistemological level, there is obviously a problem with combining reductionist with holistic approaches. In practical terms this might express itself in the evaluation questions being asked. For example, rather than asking 'what factors need to be present in order for this intervention to work?' we might ask 'how do the factors interact with each other, how do these interactions change over time, and to what extent are these amenable to intentional change?'. The underpinning theories and concrete practices of evaluations based on complexity thinking are likely to look rather different. Furthermore, if applied to the practice of evaluation rather than research more generally, Kernick's later point is perhaps more radical than is acknowledged. What would be the consequences of seeing the purpose of evaluation as supporting how 'the system will self-organize within a framework of broader policy objectives'?

The implication would be that evaluations should more often be conducted in real time and support reflexive learning and informed adaptation. Rather than seeing an intervention as a fixed sequence of activities, organized in linear form, capable of being duplicated and repeated, we see an intervention as including a process of reflection and adaptation as the characteristics of the complex systems become more apparent to practitioners. The evaluation aims in real time to understand these and support more informed adaptation by practitioners. It also provides an account of

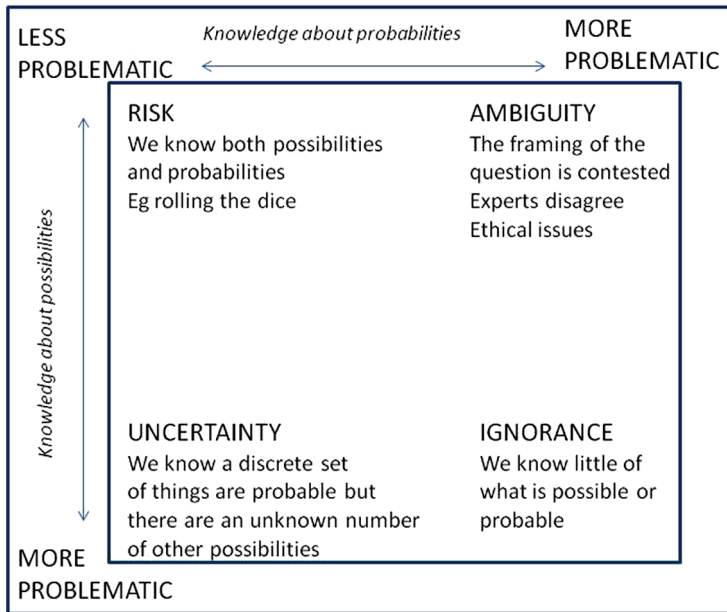


Figure 1. Stirling's matrix of unknowns

if and how effectively practitioners have adapted their activities in the light of the intended goals. They can be held to account for their intelligent adaptation rather than slavishly adhering to a set of instructions. Furthermore, the evaluation should say something about how the approach might be applied elsewhere.

Uncertainty and evaluative evidence

The knowledge produced by such real-time, embedded evaluations will explore and explain uncertainty. We have already suggested that complexity creates uncertainties which evaluations need to acknowledge and explore. In order to do this it is helpful to recognize that there are different kinds and causes of uncertainty. Stirling (2010) has helpfully constructed an 'uncertainty matrix' which suggests that our knowledge about uncertainty derives from two factors (see Figure 1). The first is our knowledge about **probabilities** (i.e. the chance of some particular thing happening) which may be unproblematic or problematic. For example, how likely it is that a roll of the dice will produce a particular number, is unproblematic. Second, there is knowledge about **possibilities** (i.e. the range of things that might happen) which might similarly be unproblematic or problematic. The range of possibilities is unproblematic where, for example, there are only two possibilities; the patient recovers or not. Where knowledge about both probabilities and possibilities are unproblematic (i.e. we know what the range of possibilities looks like and the likelihood of each possibility actually happening), Stirling argues, we can use risk assessment, optimizing models, or expert consensus. He describes this as a condition of **risk**. Where the possibilities are relatively limited but the probabilities are problematic, we can use scenarios, sensitivity testing, interval analysis and so forth. Stirling calls this a condition of **uncertainty**. Where our knowledge about the probabilities and the possibilities are both problematic we must rely upon monitoring

Table 1. Summary of differences in evaluating simple, complicated and complex interventions

Degrees of complexity and evaluation tools	
Types of intervention	Tools for evaluators
Simple interventions: <i>statistically predictable (even if individually unpredictable); information deficit; counterfactual evidence available</i>	Data on inputs, processes, outputs and outcomes, understanding (stable) preferences, cost benefit analysis, actual comparator data, ex-post or ex-ante analysis can be sufficient AIMS FOR CERTAINTY
Complicated interventions: <i>alignment of many parts is hard to achieve, 'dose' and 'frequency' hard to measure; multiple counterfactuals</i>	Logic models, modelling, simulation models, triangulating to check how the model really works, much data tacit and held by practitioners, benefits from real-time evaluation, theoretical comparator data (e.g. through Markov modelling) may be best available AIMS TO REDUCE (KNOWN) UNCERTAINTY
Complex interventions into complex environments: <i>characterized by learning, adaptation, and sensitivity to starting point. Intervention and context hard to separate</i>	Need to understand both activities and contexts, important to identify how learning and feedback happens, understand both system dynamics but also what makes change 'sticky', real-time evaluation necessary, requires a counterfactual space or matrix AIMS TO SUPPORT SELF-IMPROVING SYSTEM

and surveillance, being flexible and adaptable. This is a condition of **ignorance**. Finally, where knowledge about possibilities is problematic but probabilities less so, we can rely upon participatory deliberation, multicriteria mapping and so forth. This he describes as a condition of **ambiguity**. Whatever labels we wish to apply, risk, uncertainty, ignorance and ambiguity are clearly different from each other and reducing each requires a different approach.

The implication for evaluating complex interventions is that this offers a way of thinking about how to bring uncertainty into the evaluation. It also reminds us that different sorts of uncertainty can be understood in different ways and that the evidence for this understanding will differ. For example, where knowledge about possibilities and probabilities are both relatively unproblematic, we can use evidence from expert consensus-seeking activities such as Delphis. However, Delphis would be of no use where both are highly problematic; experts and amateurs are equally unable to deal with deep ignorance (as we have seen recently with 'economic experts'). Similarly, where the knowledge about possibilities is problematic but probabilities are unproblematic, we can use modelling or participatory deliberation to understand how people and systems might interact/be interacting in an intervention.

Complexity is not the same thing as uncertainty; it generates degrees of uncertainty but it is also associated with patterning. An evaluation based on complexity thinking should be sensitive to these uncertainties and aim to first expose them and secondly reduce them. This is discussed further below. First the article summarizes some of the evaluation tools associated with evaluating simple, complicated and complex interventions below.

Framing evaluations of complex interventions to reduce uncertainty

Elsewhere (Ling and Villalba, 2009), I have argued for an evaluation approach based firmly upon understanding the unfolding 'Contribution Stories' that those involved in delivering and

adapting interventions work with to describe their activities and anticipated effects. This approach draws upon the work of John Mayne (2008, 2010). This requires further effort to incorporate complexity thinking.

'Contribution Stories' (in the sense used here) aim to surface and outline how those involved in the intervention understand the causal pathways connecting the intervention to intended outcomes. They capture the narratives of practitioners and service users describing how their activities produce intended and unintended outcomes. This process also provides an opportunity to explore their thinking about how the different aspects of the intervention interact with each other and with other systems. From these participants' Contribution Stories more abstract Theories of Change can be developed which trace the causal pathways linking resources used to outcomes achieved. These Theories of Change will be contingent and context-dependent and should be expressed as 'mid-range theories'; not so specific that they amount to nothing more than a listing of micro-level descriptions of the causal pathway of the specific intervention but also not so abstract that it cannot be tested or informed by the evidence from the evaluation. From these two stages, evaluators can identify the key uncertainties associated with the intervention – those anticipated causal linkages for which there is limited evidence or inherent ambiguity or ignorance. Data collection and analysis would then aim to reduce these uncertainties, hopefully producing evidence that would be both relevant and timely. We suggest that this approach applies to all evaluations but it is especially helpful in relation to complex real-time evaluations. At each stage we must reflect on the consequences of complexity. These six stages are:

1. *Understanding the intervention's Theory of Change and its related uncertainties.* In emergent projects designed to accommodate the uncertainties associated with complexity, the Theory of Change will include attention to the importance of learning and adaptation. It will also identify the key dependencies upon systems and subsystems which lie outside the formal structures of the intervention. In healthcare, for example, this might include the systems shaping and changing professional and service-user identity, improvement–audit–improvement loops, dynamically evolving relationships between professional groups and so on.
2. *Collecting and analysing data focused on key uncertainties.* Methods used will be informed by Stirling's 'uncertainty matrix' outlined above. This should not only identify where the key uncertainties exist in the delivery chain but also consider what sort of uncertainty this is; ignorance or a lack of evidence; ambiguity; risk; and so forth. It may not be necessary to use the same categories as Stirling but if not, it is important to distinguish in some other way between different sources of uncertainty if we are to address these in the evaluation, particularly noting that data collection alone may not address all the key uncertainties. For example, there can be a problem when conducting an economic evaluation of a complex intervention when the client believes that completing a successful cost–benefit analysis is only a matter of collecting data on costs and benefits. More often the uncertainties arise from the boundaries and extended causal pathways with inherently unpredictable interactions that are unlikely to lead to generalizable conclusions.
3. *Identifying how reflexive learning takes place through the project and planning data collection and analysis to support this, strengthening the formative role of the evaluation.* In complex interventions there is often an overlap between the role of the external evaluation and the creation of evidence by the project itself as it learns and adapts. The external (or independent) evaluation can support this learning as part of a formative role at the same time as building a data base for its own summative evaluation. The balance between these

roles might shift significantly towards a more formative role in evaluating complex interventions. This therefore takes on more of the characteristics of a real-time evaluation.

4. *Building a portfolio of activities and costs.* The problem is not simply one of tracking changes in the cost base of the project (although this is an important-enough problem) but also that identifying boundaries around the cost base is made difficult when the success of a project may depend more on harnessing synergies from outside the intervention itself. Very importantly, a major cost in conditions of complexity is equipping projects to be adaptable and responsive to a changing environment. Essentially, part of what is being 'bought' is flexibility and, by definition, this means that some resources might not need to be used. It could be regarded as the cost of uncertainty. It is the equivalent of carrying first aid kits or reserve supplies of water on expeditions; even if they are not used they still produce a value (although a value invisible to conventional value for money studies). Real-time evaluation should reduce these costs. Indeed, this is an important justification of the costs of evaluating complex interventions.
5. *Understanding what would have happened in the absence of the intervention.* With simple interventions it is often easy to identify what would have happened had there been no intervention. Although in complex interventions it is often much harder to identify the counterfactual, nevertheless it is crucial to pose the core question in an evaluation which is 'did it make a difference?'. Answering this question must in turn involve asking 'compared with what?'. We argue that any evaluation must address the counterfactual if an evaluation judgement is to be arrived at. However, the counterfactual for a complex intervention is not a single outcome but a counterfactual space of more or less likely alternative states. This might be produced by scenarios, modelling, simulation, or even expert judgement depending upon the nature of the uncertainty.
6. *The evaluation judgement should not aim to identify attribution (what proportion of the outcomes was produced by the intervention?) but rather to clarify contribution (how reasonable is it to believe that the intervention contributes to the intended goals effectively and might there be better ways of doing this?).*

Conclusions

This short article is intended to outline a general argument explaining why evaluating complex interventions raises particular challenges and identifying the broad outline of an approach that might be suitable for conducting such evaluations. Within this, important details have not been developed. How would we in practice categorize the sorts of uncertainties identified by Stirling when conducting complex interventions? How might we build a portfolio of costs let alone price the cost of managing uncertainty? Finally we would need to think carefully about how to ensure the independence of such real-time evaluations where there must be a risk that the evaluators would become an extension of the intervention team.

Depending on the nature of the uncertainties, we face different challenges in addressing the counterfactual space. Faced with deep ignorance we may need a period of monitoring, capacity building, and hypothesis development. This is the equivalent to beginning a real-time evaluation in the very early hours and days of a humanitarian crisis, before much is known about the problem, the infrastructure, and the resources available. However, as we find out more about the possibilities and probabilities, even though complexity thinking tells us that no complex intervention is likely to produce the same outcome if repeated, it becomes possible to begin to develop a counterfactual argument. Even where we can neither duplicate the intervention nor sensibly distinguish between

intervention and control sites, we can link our counterfactual thinking to the mid-range theoretical arguments shaping the contribution story. The evaluation can collect data to explore the claims flowing from the logic of the intervention. These data can be compared with other evidence (either collected as part of the study or elsewhere from reliable sources). For example, the claim that more integrated health care services results in improved patient experience rests upon an argument of how complex inter-agency working can change relationships between service users and providers. But this complexity does not prevent us collecting patient-survey data, conducting interviews with patients and providers, comparing experiences before and after, comparing with developments elsewhere, and locating the findings within the existing body of prior knowledge. However, importantly, it warns against excessive confidence in our understanding of how sustainable, scalable and transferrable is the model based on the evidence from the evaluation.

For situations where the broad drivers of change are (relatively) certain but how they might interact is not, an alternative way to explore the counterfactual space would be to construct scenarios and ‘wind-tunnel’ the intervention in different imagined but credible futures. This would be based on the analysis of the data collected, but would seek to answer questions about scalability, transferability and sustainability. This could also be developed as an approach to *ex ante* evaluations in complex circumstances.

The approach being sketched out here clearly lacks detail and may not be sufficient. However, for this author, not only is the general argument persuasive and but also it fits with the experience of conducting such evaluations. First, those evaluating complex interventions not only cannot ignore the fact that interventions change as they unfold, but also that this adaptation is both necessary and unpredictable. Such interventions aim to reap the benefits of using the ways interventions can tip or shift systems towards significant benefits at low costs. The effects of the intervention may also be contingent upon behaviour changes that are anticipated long after the intervention itself has occurred. For example, by intervening early in children’s lives, it is often argued, their later involvement with the criminal justice system can be reduced with savings in costs and human misery. However, this always depends upon assumptions about how families, communities, criminal justice agencies, education and employment systems interact as the child grows into adulthood. There may be circumstances where these interactions overwhelm early interventions or even where the early intervention interacts to make things even worse. This can often be the case when such interventions work well as pilots but the effects diminish with scale and over time. Consequently the attractiveness of moving upstream or using joined-up working to address deep-rooted problems is likely to persist but so too will the challenge of evaluating these.

Key to dealing with this challenge is to think more carefully about the uncertainties surrounding the intervention. In highly complex interventions these can ultimately become so deep-rooted that both evaluators and implementers are in a condition of deep ignorance. Under these circumstances rapid learning cycles, sharing experiences of success and failure, and establishing monitoring arrangement might be the most appropriate way forward (e.g. delivering aid immediately following a disaster in a developing country where systems are little known, data collection challenging, and starting points obscure). In less extreme cases we might still want to move a long way from a large-scale *ex post* study aiming to nail down accountability and generate generalizable conclusions to determine future decisions.

Abandoning the search for ‘incredible certitude’ (Manski, 2007), and embracing the idea that an evaluation should aim to reduce key uncertainties, leads us to an approach which takes place in real time, is formative in nature, and aims to produce findings that are explicitly contingent. Being explicitly contingent, providing these contingent factors are adequately explained, prevents universal conclusions being drawn, but they do allow lessons to be drawn and applied elsewhere.

However, such an approach might disappoint and frustrate clients who find a lack of certitude disorienting and users of evaluation may need to have their expectations managed.

In conclusion, evaluating complex interventions involves both a challenge and an opportunity to evaluative activities. Persisting with conventional approaches and aiming for generalizability and certainty will only lead to frustration at best, and misled decisions at worst. Applying the more modest approach outlined here will support improvements based on experimentation, learning and an incremental approach to dealing with the problems of living within a complex and reflexive world.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Notes

1. For an interesting account of how real-time evaluations were delivered in Darfur see: [http://protection.unsudanig.org/data/darfur/protection/Inter-Agency%20Real%20Time%20Evaluation%20\(Jan06\).pdf](http://protection.unsudanig.org/data/darfur/protection/Inter-Agency%20Real%20Time%20Evaluation%20(Jan06).pdf) and for a more general discussion see: <http://www.odihpn.org/report.asp?id=2772> and for a guide on how to do RTE, see <http://www.alnap.org/pool/files/rteguide.pdf> (all accessed August 2011). For a more recent discussion, see: Groupe URD (2010) Real-time Evaluation of the Response to the Haiti Earthquake of 12 January 2010 (available at www.urd.org).

References

- Anderson R (2008) Editorial: new MRC guidance on evaluating complex interventions. *British Medical Journal* 337: a 1937.
- Bettis R and Prahalad CK (1995) The dominant logic: retrospective and extension. *Strategic Management Journal* 16(1): 5–14.
- Campbell N, Murray E, Emery J, Farmer A, Griffiths F, Guthrie B, Lester H, Wilson P and Kinmouth AL (2007) Designing and evaluating complex interventions to improve healthcare. *British Medical Journal* 33: 455–9.
- Giddens A (1990) *The Consequences of Modernity*. Cambridge: Polity.
- Goldstone RL (2006) The complex systems see-change in education. *The Journal of Learning Sciences* 15(1): 35–43.
- Hallam A (1998) Evaluating humanitarian assistance programmes in complex emergencies. *Relief and Rehabilitation Network, Good Practice Review* 7. London: ODI. URL: <http://reliefweb.int/sites/reliefweb.int/files/resources/166585C863B5E414C1256D500058A950-rrn-evaluate-sep98.pdf>
- Hawe P, Shiell A and Riley T (2004) Complex interventions: how ‘out of control’ can a randomized controlled trial be? *British Medical Journal* 328: 1561–3.
- Institute of Medicine (2001) *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academies Press.
- Kernick D (2006) Wanted – new methodologies for health service research: is complexity theory the answer? *Family Practice* 23(3): 385–90.
- Ling T and Villalba L (eds) (2009) *The Performance Audit Handbook*. Cambridge: RAND Europe.
- Ling T, Bardsley M, Adams J, Lewis R and Roland M (2010a) Evaluation of UK integrated care pilots: research protocol. *International Journal of Integrated Care* 10, 27 September. URL: <http://www.ijic.org/index.php/ijic/article/download/513/1069>
- Ling T, Soper B, Buxton M, Hanney S, Scoggins A and Steel N (2010b) *An Evaluation of the Health Foundation’s Engaging with Quality Initiative: Final Report*. London: The Health Foundation.
- Litaker D, Tomolo A, Liberatore V, Stange KC and Aron D (2006) Using complexity theory to build interventions that improve health care delivery in primary care. *Journal of General Internal Medicine* 21: 30–4.

- Manski C (2007) *Identification for Prediction and Decision*. Cambridge, MA: Harvard University Press, 7–8.
Quoted in Manski C (2011) Policy analysis with incredible certitude. Paper presented at the Institute for Fiscal Studies, February 2011.
- Martin C and Sturmberg J (2009) Perturbing ongoing conversations about systems and complexity in health services and systems. *Journal of Evaluation in Clinical Practice* 15(3): 549–52.
- Mayne J (2008) Contribution analysis: an approach to exploring cause and effect. ILAC Brief 16. URL (consulted August 2011): http://www.cgiar-ilac.org/files/publications/briefs/ILAC_Brief16_Contribution_Analysis.pdf
- Mayne J (2010) Contribution analysis: addressing cause and effect in simple and complex settings. In: Schwartz R, Forss K and Marra M (eds) *Evaluating the Complex*. Piscataway, NJ: Transaction Publishers.
- Medical Research Council (2008) *Developing and Evaluating Complex Interventions: New Guidance*. London: Medical Research Council.
- Pawson R, Greenhalgh T, Harvey G and Walshe K (2005) Realist review – a new method of systematic review designed for complex policy interventions. *Journal of Health Services Research and Policy* 10(1): 21–34.
- Pederson LH and Reiper O (2008) Is realist evaluation a realistic approach for complex reforms? *Evaluation* 14(3): 271–93.
- Rogers P (2008) Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation* 14(1): 29–48.
- Sampson A (2007) Developing robust approaches to evaluating social programmes. *Evaluation* 13(3): 477–93.
- Sanderson I (2000) Evaluation in complex policy systems. *Evaluation* 6(4): 433–54.
- Sanderson I (2009) Intelligent policy making for a complex world: pragmatism, evidence and learning. *Political Studies* 57: 699–719.
- Shiell A, Hawe P and Gold L (2008) Complex interventions or complex systems? Implications for health economic evaluation. *British Medical Journal* 336: 1281–3.
- Soper B, Buxton M, Hanney S, Oortwijn W, Scoggins A, Steel N and Ling T (2008) Developing the protocol for the evaluation of the health foundation's 'engaging with quality initiative' – an emergent approach. *Implementation Science* 3: 46.
- Stirling A (2010) Keep it complex. *Nature* 468: 1029–31.
- Stollar SA, Poth RL, Curtis MJ and Cohen RM (2006) Collaborative strategic planning as illustration of the principles of systems change. *School Psychology Review* 35(2): 181–97.
- Van Stolk C, Ling T, Reding A and Bassford M (2011) *Monitoring and Evaluation in Stabilisation Interventions: Reviewing the State of the Art and Suggesting Ways Forward*. Cambridge: RAND Europe.
- Weaver W (1948) Science and complexity. *American Scientist* 36: 536.

Tom Ling is Director for Evaluation at RAND Europe. He has 20 years experience in evaluation. Recent publications include *Performance Audit: Contributing to Accountability in Democratic Government*, *Performance Audit Handbook* and *The Evidence Book*. He is Professor of Public Policy (Emeritus) at Anglia Ruskin University, and Honorary Senior Visiting Research Fellow, University of Cambridge. Please address correspondence to: RAND Europe, Westbrook Centre, Milton Rd, Cambridge, CB4 1YG, UK. [email: tling@rand.org]