The Practice and the Theory

Lessons from the Application of a Theories of Change Approach

MHAIRI MACKENZIE AND AVRIL BLAMEY

University of Glasgow, UK

Theory-based evaluation approaches are becoming increasingly popular in the evaluation of comprehensive community initiatives. Such an approach, Theory of Change (ToC), has been used in the external evaluations of two of four Scottish Health Demonstration projects. This article provides an empirically based analysis of the issues raised in the process of articulating a project's ToC. It then considers the value of the approach in relation to sharpening planning, providing formative feedback, improving performance management, guiding internal and external evaluation, judging impact and reducing problems of attribution. A conclusion is reached that, whilst such approaches are by no means a panacea, they offer one useful framework within which to negotiate a range of evaluation practice decisions.

KEYWORDS: child health improvement; complex community interventions; heart health improvement; Scottish Health Demonstration projects; theory-based evaluation

Introduction

Theory-based evaluation approaches such as Theory of Change (ToC) (Fulbright-Anderson et al., 1998) or Realistic Evaluation (Pawson and Tilley, 1997) are becoming increasingly popular in the evaluation of social programmes and comprehensive community initiatives. Their rise in popularity stems from the proliferation of the commissioning of complex interventions and the hope that such approaches can overcome the limitations of previous social programme evaluations (Wimbush and Watson, 2000). Rather than simply judging complex community interventions in their entirety as successes or failures (Susser, 1995) there is a recognized need to uncover what aspects of an intervention have or have not been successful with which particular subgroups (Pawson and Tilley, 1997).

The ToC approach has primarily arisen from the work undertaken by the Aspen Institute Roundtable Initiative (Fulbright-Anderson et al., 1998). Connell and Kubisch (1998), two key proponents of the approach, claim that it has three key strengths in relation to previous approaches to social programme evaluation.

They suggest that the approach can sharpen programme planning, can facilitate decisions concerning the prioritization of evaluation questions and methods and can reduce the problems associated with causal attribution that commonly plague the evaluations of complex interventions. These three goals are common to many evaluators.

This article will focus on two main issues: first, it will discuss some of the practical considerations in working with a project to prospectively articulate its ToC; and second, it will critically assess the degree to which Connell and Kubisch's claims are substantiated through the experience of using the approach. Lessons will be drawn from the use of the approach within two single-site complex community interventions.

The evaluations that the article will draw upon are the independently commissioned evaluations of two Scottish Health Demonstration projects (SHDPs). The two projects covered the areas of heart disease prevention (Have a Heart, Paisley [HaHP]) and child health (Starting Well [SW]), and were both funded, for an initial three-year period, by the Scottish Executive (SE) to the sum of £6 million and £3 million respectively. In addition to improving health outcomes in these two topic areas through partnership working, the projects were tasked with reducing health inequalities and fully engaging and involving the community in all aspects of the programmes. The interventions were given some limited resources to aid internal evaluation (e.g. to fund a relatively junior research post) and the Chief Scientist Office (one of the Scottish Government's research commissioning bodies) of the SE commissioned independent evaluations from the University of Glasgow (and key partners). Whilst a ToC approach was used within these projects, it formed only one aspect of a wider evaluation that had quasi-experimental pre and post intervention surveys, context analyses and a range of pre-specified qualitative investigations. (The benefits of combining a ToC approach with traditional experimental methods are discussed by Weitzman et al., 2002.) The independent evaluations began work after the actual programmes had been launched.

Prospectively Articulating a ToC: Moving from the Aspen Theory to Evaluation Practice

This section of the article discusses the framework offered by the Aspen Institute as a means of articulating a ToC and assesses its usefulness in producing a well-specified summary of the links between the contexts, planned activities and predicted outcomes of an intervention. The implications for the role of the evaluator are also discussed.

When to Surface a ToC

Connell and Kubisch start with the acknowledgement that the ToC articulation process is not a straightforward one: 'In our experience, surfacing and articulating a ToC through a collective and collaborative process is as fraught with difficulties as it is full of promise' (Connell and Kubisch, 1998: 21).

It should not, therefore, be entered into lightly nor seen as a short cut to developing an evaluative plan. Notwithstanding the iterations, political negotiations and conflict resolutions inherent in the process, a set of steps is outlined that, if followed sensitively, will result in a project's ToC.

These steps are as follows: identification of the long-term outcomes that the initiative seeks to achieve; identification of the interim outcomes and contextual features that will be required to meet these longer-term outcomes; specification of the activities that will be put into place and the contextual requirements to realize these interim outcomes; and an explicit recognition of the resources that will be required to turn these goals into reality.

The first implication of the above is that the ideal ToC is undertaken in the planning phase of an initiative. This is the point at which there is arguably more opportunity for the acknowledgement of gaps between aspirations and the reality of implementation and where individual stakeholders are more open to considering competing analyses of the problem and its solution.

The reality for the SHDPs, as with the vast majority of UK government-funded initiatives, is that external evaluation teams and project planners rarely have this window of opportunity. The projects themselves entered into a bidding process and won funding on the basis of their plans but the time for submission of these plans was limited and hardly conducive to wide consultation. The proposals for the external evaluation were (in the early stages at least) developed at some distance from the projects themselves.

This confirms the analysis of Sullivan et al. (2002) in their description of the limitations of a ToC approach in relation to the National Evaluation of Health Action Zones,

Even though it is now common in UK public policy for an evaluation requirement to be part of the process of policy implementation, the time taken to invite tenders and negotiate the brief means that it will be rare for evaluation and programme to develop contemporaneously. (2002: 208)

In addition, whilst a ToC approach formed one strand of the overall external evaluations, there was a recognition that proposals should include a reasonably clearly delineated quantitative component in order to be successful in securing funding from central government. It is clear then that there are barriers to undertaking the approach at a time when stakeholders may be most receptive.

The Role of the Evaluator in Articulating and Testing Theories

Connell and Kubisch also acknowledge that the approach raises some difficult issues in relation to the role of the evaluator and, in particular, the extent to which s/he can act as an impartial and objective researcher employing technical skills alone. For example, where conflicting project rationales emerge they say that, 'Resolving the challenges that these multiple theories pose is a political as well as scientific process' (1998: 30).

It is useful then to reflect on what some of these challenges look like in practice and to consider how the ToC approach contributes to blurring the role between evaluator and implementer. (These considerations are not offered as a reason to

avoid the approach; on the contrary, they are offered as a means of acknowledging the complex contribution the approach can make.)

First, there is the question of who initiates the process of articulating the ToC. Herbert and Anderson (1998), for example, describe the role of the evaluator in this respect as falling into two categories. The evaluator might instigate the process by producing a tentative model that stakeholders would discuss and negotiate *or* they might focus on summarizing, critiquing and iterating the views of stakeholders building on existing documentation and plans.

Arguably, the construction of the role of the evaluator will be influenced greatly by the approach taken here. Within the SHDPs, initial ToCs were developed using the latter approach. It is tentatively suggested that the second approach is the more practical where ToCs are not being explicitly surfaced in the earliest planning phases of an initiative.

Second, for proponents of theory-based evaluations there is a central debate about the types of skills that an evaluator brings to the project table and, in particular, the extent to which domain knowledge is required. Brown suggests that domain knowledge, whilst not essential, is highly desirable: 'Our experience suggests that simply being knowledgeable about a field can help the evaluator probe the assumptions . . . build credibility with the sites, and accelerate the process of creating a framework that receives the support of all stakeholders' (1998: 107).

Domain knowledge, however, throws up its own challenges. Within the evaluation of one of the SHDPs, the evaluator in surfacing the views of one stakeholder uncovered the intention of one agency within a multi-agency partnership to set up a screening programme that had previously been shown to be of limited effectiveness in such an environment. This raised both a logistical and an ethical dilemma that was only partly reduced by providing direction of where to *seek* reliable advice on designing a more robust intervention rather than actually *providing* that advice and becoming involved in delivery. Day-to-day involvement with the two projects, however, threw up a whole range of interactions where views on delivery were sought and given by the evaluators and where it was more difficult to disentangle evaluation and implementation advice. For many evaluators this may raise serious questions about the objectivity of the evaluation and the possibility of generalizing from evaluation findings; for others, programme improvement is an explicit goal for the evaluator (Donaldson and Gooler, 2003). This is a debate that is well rehearsed elsewhere (Pawson and Tilley, 1997).

A related difficulty for both evaluators and implementers to grapple with is that close relationships developing over the life of an evaluation can lead to opposing views on the extent to which the evaluator should adopt a *critically* supportive role rather than simply a supportive role. The approach requires trust to develop and be sustained. Yet, at the same time, evaluators have a responsibility to feed back formative findings in a manner that might assist both the individual project to improve its delivery and funders to make decisions about sustaining and extending project funding (or indeed, rolling out project lessons to other geographical areas).

In addition to the potential for a blurring of roles between evaluator and project staff there is a further confusion of roles between evaluation and performance management. The differences and similarities between these two fields are discussed in detail by Blalock (1999). She indicates that the two approaches are complementary, and that there is a great deal to be achieved in integrating the two activities. Through the ToC process with the SHDPs there grew a recognition on the part of the funding body that technical assistance in performance monitoring was required and an independent consultant was commissioned to provide some short-term advice. To a varying degree across the two projects this consultant worked in tandem with the evaluator and this resulted in some confusion about the distinction between the two activities.

Whilst none of the problems discussed in this section are unique to a ToC approach, the required proximity of the evaluator to the evaluand undoubtedly exacerbates the potential for roles to be misunderstood.

When is a ToC Fully Developed?

In the next section of this article, which looks at the claimed benefits of the approach, there is a detailed discussion of the implications of the degree of specificity required in order to test the assumptions and impact of a project's ToC. In this section the need for clarity about what a ToC looks like in practice is raised.

Connell and Kubisch (1998) provide two examples of ToCs but acknowledge that these are 'at an admittedly general level'. Whilst the examples provided give a good sense of the sheer complexity of the interventions described and of the overall predicted change pathways, they don't provide a sense of what a final and usable version of these might look like. This is not a pedantic point but a practical issue that makes it difficult for evaluators to know when to stop digging for detail. Similarly there is little advice on the tools to be used in working with groups. Cole (1999) attempts to identify techniques and strategies for constructing theories using examples from public health to fill this gap.

The need to know when to draw a line around a project's initial ToC is important in a temporal sense and in terms of what the Aspen Institute might refer to as a vertical complexity sense. Thinking of the temporal sense, first, it is important to reflect on the fact that the ToC produced for and with the SHDPs represented the first year of the projects' lives. The projects themselves, however, continued over time to refine their activities and predicted outcomes and so the degree to which there is ever a definitive representation of a project's plan is questionable. (This has implications for the number of times a ToC is revisited and deviations explored.)

In terms of vertical complexity, work with HaHP in particular illustrated the need to surface ToCs not only at a strategic project management level but also at an operational level with individuals delivering individual strands of the intervention. Without some of the detail provided by the operational level staff the strategic ToC lacked the ability to be interrogated in relation to its practicality and measurability as a plan; without the strategic level theory, the operational theories were disjointed and lacking in complexity.

Connell and Kubisch (1998) recognize the tension between expectations of

great detail at the beginning of the process and later opportunities to capture the unexpected outcomes associated with complex inter-relationships,

Imposing strict standards of theory articulation too early in the process can undermine participation and stifle the dynamic nature of the CCI [comprehensive community intervention] enterprise. At the same time ... leaving the CCI's ToC ambiguous permits, and indeed, encourages, various stakeholders to project their own preferences about activities and outcomes onto the initiative. (1998: 30)

In practical terms, however, this has the tendency to leave the evaluator between a rock and a hard place, grappling for guidance and a foot-hold.

A final issue concerns the capturing of complexity within what is essentially a linear model. Although many of the contributors to the Aspen Institute approach recognize the need for theories of change to reflect the complex inter-relationships between activities, context and outcomes, within and between different strands of the one intervention, the steps required to elicit the ToC are suggestive of a relatively linear view of the world and its organizations. Whilst this provides a useful short-hand for project planning it can be argued that it is potentially a constraining approach that actually impedes thinking in a synergistic and partnership manner (Barnes et al., 2003). Within SW, for example, from a health perspective, improved health outcomes might be viewed as the ultimate goal with improved family functioning as an intermediate objective, whereas the converse might be true for those operating within social services.

The Utility of a Good ToC

The previous section explored some of the difficulties that an evaluator has in putting into practice the ToC framework offered by the Aspen Institute. The current section begins by considering further difficulties in articulating a 'good' ToC (with examples from the SHDPs) before explicitly assessing the claims put forward by Connell and Kubisch (1998) for its usefulness as an approach.

What is a 'Good' ToC?

According to Connell and Kubisch, a good ToC is one that is plausible, do-able and testable. Plausibility refers to the extent to which the planned activities are linked, through an existing evidence base or at least an inherent logic, to the problems identified and the outcomes that they aim to achieve. Do-able relates to the degree to which the activities are deliverable within the timescales, context and resources available to the project. Testable relates to the extent that the theory is well enough specified to allow verification of progress in delivering the plan and subsequently, the measurement of the intended outcomes.

Plausibility

The overall ToCs initially uncovered with the demonstration projects were *on the whole* plausible in the sense that they attempted to tackle the widely acknowledged key determinants of poor health in their respective topic areas. (See Figures 1 and 2, Box 1 and Table 1 for the initial representations of HaHP ToC. These

Figure 1. Have a Heart, Paisley – Simplistic Theory of Change

depict a simplistic theory, a more detailed overall theory and a specific theory and rationale for one of the 17 individual work strands.) For example, HaHP through primary and secondary prevention set out to reduce and prevent key clinical and behavioural risk factors associated with CHD (coronary heart disease), whilst SW attempted, through enhanced home-visiting and improved access to services, to tackle lack of social and parenting support for families of newborn children living in socioeconomically deprived and difficult circumstances. Once evaluators uncovered more detailed operational plans, however, in *some* areas the plausibility in relation to actual delivery looked less ‘convincing’. In some project areas of HaHP the delivery of the interventions did not necessarily follow existing best practice guides. For example, some of the physical activity programmes were predominantly facility-based yet active living approaches have been shown to be equally successful and less expensive (Dunn et al., 1999). Such approaches may have been more appropriate given the limited facilities available locally.

Do-ability

The theories of change uncovered for the projects were do-able, in terms of finance, but were less feasible in relation to time. This was in part due to the early political imperative, voiced by commissioners and responded to by implementers, to establish activity on the ground as soon as possible and this concurs with lessons drawn by other evaluators (Sullivan et al., 2002; Walker, 2001). This led

PA = Physical Activity
 PC = Primary Care
 SC = Secondary Care
 HPHS = Health Promoting Health Service

158

Figure 2. Have a Heart, Paisley – Overall Theory of Change

Box 1. Work Theme: Rationale for Improved Secondary Care (Heart Renewal Project)

Context

The plans for the Heart Renewal programme (there are 16 other project strands of a similar nature) are based in the context that currently only those who have suffered a myocardial infarction (MI), or are receiving surgical interventions, will be offered access to a formal rehabilitation programme. It is hoped that the extension of facilities and the development of a new wider, menu driven approach will increase the number offered rehabilitation. (This should include those with a step change in CHD or with angina rather than simply those post MI or post surgery.) It should also increase adherence to the programmes taken up which should then lead to improved quality of life and decreased secondary events. There are also plans to increase links to and extend Phase Four (community based continuity classes for those discharged from the hospital setting). While there are many rehabilitation programmes operating across Scotland the quality of such services can vary greatly and most are only offered to a limited range of patients who have presented with acute events.

Rationale/Assumptions

There is a range of evidence that indicates that participation in rehabilitation post MI can significantly reduce secondary events and improve quality of life and that participation in regular physical activity can reduce or ameliorate CHD risk factors. Although a range of relatively positive evidence exists, the results of meta-analysis and overview studies are believed by some clinicians to require further RCT evidence to fully demonstrate the potential impact of such services. An RCT, however, is not being progressed in this instance due to ethical concerns over withholding potentially beneficial services and pragmatic difficulties with conducting an RCT.

Degree of specificity

The plan is well specified. A few gaps in relation to deadlines and one short-term outcome require further specification after baselines are collected.

to some project strands being initially poorly thought through or failing to be based on actual needs assessments and available baseline information. Given the very short three-year timescale to impact on CHD and child health outcomes, it is not surprising that the projects found it difficult to articulate what could be achieved that would be acceptable to stakeholders and would be sensible intermediate steps on the way to long-term improvements. The project teams were, for example, still recruiting staff up to and after the point that they were articulating their ToC. Their expectations, therefore, regarding the skill levels of the staff they could recruit and the time it would take for staff to 'hit the ground running' were often too ambitious. The bureaucratic structures of the different organizations and skill shortages in key areas such as health visiting or evaluation meant that many of the early expected outputs were overly ambitious. Positive attempts were, however, made by SW to overcome this recognized skill shortage prior to implementation of the project; for example, the intervention was prospectively designed to test out skill mix solutions such as the employment and training of lay health workers. In relation to HaHP the internal evaluation

Table 1. Have a Heart – Specific Theory and Rationale for *Improved Secondary Care* Work Strand

Work Theme	Improved Secondary Care: Heart Renewal Programme
Implementation	<p>Employ staff by October 2001 (2 nurses, dietician, physio, exercise scientist, p/t secretary, sessional psychologist)</p> <p>Establish links and agree and write protocols (local clinical agreements) with other key professionals by July 2001</p> <p>Design and implement a menu driven rehab programme by November 2001</p> <p>Conduct baseline audit of potential target groups (those with step change in CHD) by December 2001</p> <p>Redevelop hospital facilities by October 2001</p> <p>Design and implement official phase IV programmes in community by January 2002</p> <p>Implement hand held record system by January 2002</p> <p>Conduct an audit of access to service by deprivation category</p> <p>Ensure implementation of clinical guidelines across primary and secondary care on an ongoing basis</p>
Outputs	<p>90% of eligible patients with step change in CHD status offered menu based phase III rehab</p> <p>70% uptake of some aspect of menu based phase III rehab</p> <p>Phase IV opportunities available in all four localities by January 2002</p> <p>90% of patients completing phase III offered phase IV by 2002</p> <p>70% uptake of phase IV by those offered it after completing phase III</p> <p>Audit conducted during 2002–3</p>
Short-term outcomes	<p>80% of those taking up aspects of menu based phase III completing respective programmes (>75% attendance)</p> <p>50% of those taking up aspects of menu based phase III adhering to behaviour change at one year</p> <p>50% of those taking up phase IV from phase III adhering to behaviour change targets after one year of phase IV</p> <p>Significantly improved risk factor (RF) control across all key groups and RFs compared to their own and historical baselines (exact targets for % change in those post MI/surgery and with step change for each RF to be made after baseline complete)</p> <p>Improved access and adherence to service by excluded and deprived groups relevant to baselines</p> <p>Adherence rates for those from excluded groups to move towards those for non-excluded patients</p> <p>Improved perceptions in quality, relevance and impact of service to patients</p> <p>Improvements in QOL and SF36 scores compared to baseline where measures available</p>
Long-term outcomes	<p>The prevention of secondary events, reduced morbidity and mortality and improved quality of life post event (specific targets for % reduction from current baseline to be made after completion of baseline)</p>

post was vacant for a substantial amount of time due to a shortage of skilled applicants. This had a knock-on effect for internal monitoring and formative feedback. In terms of do-ability then the ToCs were probably over-ambitious at least in relation to skills, timescales and overall capacity.

Testability

The area with which the projects had the greatest difficulty, in relation to articulating a good ToC, was that of testability. Both projects had substantial difficulty in articulating outcomes that covered the expectations of a range of stakeholders (internal and external) and that were specified to a degree that would allow measurement of progress. This was particularly the case in terms of possible quantifiable measures and in relation to specifying the magnitude of change expected. Even once measurable outcomes were identified there was enormous difficulty and/or reluctance in prioritizing these into a manageable number of key overriding outcomes that the projects would be accountable for delivering. It took, in fact, almost 18 months of negotiation between the commissioners, evaluation team and implementation teams and a range of activities, including the ToC articulation process and additional support from a performance management trainer, to establish such a set of outcome measures. Some of the reasons for such difficulties may relate to the lack of an evaluation culture or traditional evidence-based approach within statutory agencies. For example, the local authorities may have seen themselves as accountable to elected members and local residents more than to a medically orientated evidence approach. In addition to cultural issues, there were issues about a lack of confidence, knowledge and skills in relation to identifying indicators. There may have been a blame culture in many of these organizations and so asking implementers to prospectively set targets (particularly challenging targets) may have been viewed as creating a stick to 'break their backs with'. Lewis (2001) comments on how targets can be used differently when set for either performance management or evaluation or by different organizations, 'targets can either be objectives which are ideals which one does not necessarily have to achieve, so are not performance measures in themselves, or targets are like dartboards and have to be hit – and to which funding is often attached' (2001: 388–9).

Some of the limitations to developing a good ToC in all of the above senses relate to a very real lack of accessible secondary baseline data or benchmarking information from other similar services that could easily be used at a local level. For example, the national monitoring survey for CHD in Scotland (Shaw et al., 2000) did not allow data to be disaggregated to individual local authority levels, let alone the level of towns, whilst local Health Board surveys achieved generally poor response rates or had limited sample sizes due to financial constraints. There were, therefore, no reliable and valid data with regard to key issues such as smoking prevalence or incidence of angina that could easily be used by HaHP.

These barriers to developing a 'good' ToC have significant implications for the degree to which the ToC approach helped the demonstration projects to sharpen project planning, facilitate the focus of their evaluations and reduce potential difficulties in attribution.

Addressing the Problems of Causal Attribution

In relation to attribution, proponents of the ToC approach argue that 'Articulating a ToC at the outset and gaining agreement on it by all stakeholders reduces, but does not eliminate problems associated with casual attribution of impact' (Connell and Kubisch, 1998: 18).

This claim is based on the assumption that the information detailed in a prospectively specified plan can be used as an explanation for subsequent changes in predicted outcomes. The premise is that, provided activities and outcomes pre-specified in the plan have been agreed by a range of relevant stakeholders as being plausible, and that the plan has been delivered as agreed and within the expected contexts, then attribution claims should be strengthened. The way that attribution is improved, therefore, may be independent of any secondary objective measures but simply through the fact that the project has delivered exactly what stakeholders agreed would be indicative of project success. An example of this from SW is that stakeholders would have accepted that, if an intensive home-visiting service had been delivered to all new parents in the interventions site, and these parents reported through pre and post measurement with validated scales that they had increased levels of confidence in their parenting role, this would subsequently lead to increased educational attainment amongst the children in these families. The difference between attribution in this context and 'gold standard' randomized controlled trials is acknowledged by Connell and Kubisch:

Although this strategy cannot eliminate all alternative explanations for a particular outcome, it aligns the major actors in the initiative with a standard of evidence that will be convincing to them. Clearly this will not be as powerful as evidence resulting from randomly assigned control and treatment groups, but as has been noted elsewhere, random assignment of communities is not a feasible avenue of evaluation from CCIs. (1998: 16)

A further way, however, that attribution claims may be justified is that the process information captured from the ToC can be used in conjunction with objective secondary or primary outcome data to gauge whether the resulting change is explicable via the dose and exposure of target groups to the actual interventions. Confidence in attributing outcome changes to the intervention activities is increased when the evidence base for the activities is strong and when the magnitude of change is as predicted. If, for example, within HaHP the stakeholders had decided to judge the success of the project by the degree to which the process information explained a 2 percent reduction in smoking achieved in the intervention site after five years, it would be vital to know the range and precise numbers of opportunities for cessation provided (e.g. phone lines, pharmacy support, group and one-to-one interventions) and the number of patients who accessed these. Only from existing knowledge of the effectiveness of such services, or from additional local primary data about adherence to therapies and sustained quit rates combined with knowledge about other contextual changes in the area (e.g. national media campaigns or increased number of smoking policies), would the activities be judged with any confidence to have been responsible for such a magnitude of change.

Acknowledging that there are different ways to interpret a ToC's potential to improve attribution is important as the degree of specificity required and the range of stakeholders consulted may vary according to the approach adopted. Linking the process information to secondary or alternative primary data sources (such as quasi-experimental surveys) requires a greater degree of specificity and detail of both the evidence base and process information of planned activities, and the actual magnitude and timing of expected change within a ToC.

In relation to the two demonstration projects, the detailed ToCs that were articulated were quite different. The HaHP approach was more linear and aspects of it were specified to a greater extent than for SW. This was probably due in some part to the nature of the available evidence about the two topic areas. CHD has fairly well recognized clinical and behavioural risk factors that, although influenced by social factors such as poverty and educational attainment, can to some degree be isolated from these wider determinants and can be targeted by a range of drug therapies, counselling and behavioural approaches, policies and structural changes. SW in relation to child health, on the other hand, was dealing with a less well-developed evidence base and was trying to impact on a range of childhood health conditions and outcomes. SW represented the first time that this approach to the delivery of an enhanced home-visiting support service had been tested in Scotland.

The different linearity and specificity might also have been due to the types of stakeholders involved. There were probably a greater number and dominance of clinicians in the CHD project as a result of the emphasis on both primary and secondary prevention interventions that have been shown to influence heart disease. The HaHP projects included rehabilitation, the establishment of a disease register, the use of drug therapies and the involvement of primary and acute care teams as well as a wide range of community-based activities and implementers. Given that both projects had a quasi-experimental design element to them, it will be interesting to see if they differ in their use of the ToC in relation to attribution in the future.

The range of stakeholders might also determine which of the interpretations of attribution dominates theory articulation. Conversely, however, if evaluators or implementers select one or other of these approaches prior to wide consultation this might limit the range of stakeholders actually involved in theory articulation or indicator selection. There is, in other words, a potentially interesting interplay between the constituency of stakeholders and the interpretation of attribution. Attribution is therefore, like evidence, a contested concept.¹

For both approaches it is important that the changes from the initial to the final theory articulation process are captured so that explanations are based on what has actually been delivered.

Sharpening Planning and Implementation

Although the articulation of the ToC was a slow and difficult process for the evaluators and many of the implementers, there is no doubt that it has led directly to improved planning within the projects themselves. The approach led to specific improvements in relation to surfacing discrepancies in stakeholders' priorities

and subsequently improving the focus of activities. In relation to SW, an example of this was that three key stakeholders from within the same organization identified three different outcomes as the key priority for the project in the early stages of the articulation process: one identified improved cognitive functioning as the most important outcome; the second, improved health outcomes; and the third, improved social functioning.

As discussed earlier the Aspen Institute papers do not specify tools for use in uncovering a ToC and so the evaluators borrowed tools from performance management such as logic models (Rush and Ogborne, 1991) and balanced scorecards (Accounts Commission for Scotland, 1998; Kaplan and Norton, 1996). These tools were used to uncover the ToC and identify key outputs and outcomes across long- and short-term timescales. The use of such tools, though time-consuming, ensured a consistent and detailed approach to plans across the different agencies and projects.

As discussed, the latter stage of the ToC process where indicators of success and monitoring methods were being selected was particularly difficult for the staff of some agencies such as the local authorities and the primary care trust. Again this is likely to have been due to a lack of an evaluation culture or differing approaches to service planning across these agencies.

Another important issue that was raised as a result of the ToC process was that of capturing the synergy and interaction that result from the simultaneous delivery of a range of linked interventions. This issue is key for both planning and evaluation. In relation to HaHP this relates to whether the project was trying simply to deliver 17 separate but quite complex interventions or whether the whole was more than the sum of these 17 parts. In many ways this project was not actually delivering many truly innovative interventions but was utilizing evidence-based programmes from elsewhere. The difference from other areas of Scotland was that these interventions were being delivered simultaneously and on a larger scale. In this sense, the project was in fact testing the impact of the synergy between the activities, the additionality of joint delivery, and whether or not they could saturate the intervention town with a high dose of messages about and opportunities for change. There were similarly substantial discussions around roles and responsibilities for delivering cross-cutting outcomes such as partnership working, tackling inequalities and community involvement. This led to an acknowledgement of the need to recognize these issues as both outcomes and processes and the necessity to integrate the responsibility for these issues across programmes and find means of measuring change in these areas.

The usefulness of the ToC process in relation to sharpening planning was frequently commented on by the implementation teams. Some of the project staff subsequently decided to utilize the approach for future planning exercises with new projects. Overall then, despite being difficult and time-consuming, the process was welcomed and viewed positively.

Part of the success of the approach in sharpening planning resulted from the fact that the ToC process and reports acted as a catalyst which made commissioners request greater specification of plans in areas where detail was still lacking. To help achieve this, they offered projects additional support from an

independent performance-management trainer. This support built on the ToC process but used slightly different tools and methods. Similarly the process encouraged the SE to pull together wider groups of national-level stakeholders and to focus on what the expectations of these people were in relation to all four of the national demonstration projects (two of which are SW and HaHP) and particularly in relation to common cross-cutting themes such as partnership working, community involvement and tackling inequalities. A third response from the SE was for them to request and supply further specification of expectations of, and outcomes from, the internal evaluation process.

The increasing imperative on projects for improvements in performance management and accountability raises issues about whether commissioners are clear about their own expectations from complex social programmes. In response to this issue the evaluators conducted focus groups with the SE in an attempt to elicit commissioners' ToCs. This process indicated that those at a commissioning level had learned lessons about the timing and process of commissioning, and the support required for such projects. Indeed, at the time of writing the projects had received funding for an additional year with subsequent funding contingent on the production of detailed plans for sustaining change and implementing lessons learned.

Improving the Evaluation Focus

Connell and Kubisch's (1998) final claim for the use of the ToC is its ability to facilitate measurement and data collection for evaluation. The ToC, they believe, should provide a clear route map of an initiative allowing for the selection of sensible evaluation priorities and guidance on when to measure these.

The ToC produced for the demonstration projects did indeed aid the process of identifying key areas of potential focus for both internal and external evaluation. An example of this for HaHP was an agreement that the external evaluation should focus on areas related to the whole project and cross-cutting themes, rather than the individual project strands. In addition, as a result of the information gathered through the ToC process, the independent evaluation was rewritten to allow substantial triangulation of data from a range of qualitative and newly designed smaller-scale surveys that would maximize learning for other key areas of the programme rather than population impact alone. In SW there was an agreement that a key part of the qualitative work would focus on the important relationship between health visitors and the lay workers as this was identified as a vital mechanism in the delivery of the enhanced home-visiting and parenting support programme.

Agreement on these areas of focus meant that there was opportunity for increased clarity on the subsequent focus of internal evaluation and monitoring. (Again the evaluators used tools from performance management such as an adapted version of a balance scorecard to help in this process.) This increased clarity was particularly true for SW. In some instances, however, it was difficult to encourage clarity and there were some conflicts over the balance between focusing on areas of innovation to maximize learning and on areas with a strong evidence base to highlight progress and success. With HaHP it took a long time

to identify and agree a limited number of key outcomes that would become the focus for the internal evaluation and the subsequent links with the external evaluation. It would seem, therefore, that the ToC did raise issues of conflict with regard to evaluation; however, substantial additional time, work and tools were required to resolve these conflicts.

Concluding Comments

This article has explored the benefits and limitations of a ToC approach to the evaluation of complex single-site initiatives by focusing on two issues: the practicality of articulating a ToC; and the utility of the resulting ToC for the project, its commissioners and external evaluators.

Learning from the demonstration projects, thus far, supports the utility of the ToC process. However, it is argued that the three key benefits identified by Connell and Kubisch (1998) are not unique to this evaluation approach; nor can they be achieved without overcoming some of the many practical barriers that exist to uncovering a good ToC.

The process is exceptionally time-consuming and requires dedicated evaluation time from individuals who have good evaluation field experience and preferably have domain knowledge. A range of tools are required that facilitate the articulation of detailed plans from staff at different levels within organizations. Various methods should be employed (documentary review, group work and interviews) to ensure the involvement of a wide range of stakeholders and ownership of the uncovered ToC. Theories should ideally be articulated at both strategic and operational levels and the links between the two need to be made explicit.

The common problems faced by implementers with regard to accessing valid local data need to be addressed and alternative and innovative approaches to establishing baselines are required. Substantial work is needed to encourage appropriate monitoring procedures within projects. These procedures can represent a significant resource both in time and money and their value needs to be recognized by implementers at a strategic and operational level (Accounts Commission for Scotland, 2000).

There are also some limitations to the ToC approach. To some degree it can still lead to very linear approaches to planning and evaluation. Such linearity may miss or mask some of the very complex interactions within and between projects or across target groups or areas. Such linearity is also a distortion of the way in which most organizations develop and learn from strategy and implementation (Sanderson, 2000).

The approach may also be unable to uncover unexpected outcomes or synergies. The ability of a ToC, for example, to facilitate the measurement of cross-cutting issues such as saturation and synergy will rely on sound internal monitoring and creative use of a range of evaluation methodologies. Often the research methodologies available and the skills and procedures for monitoring that exist are not sensitive or responsive enough to the complexity that exists (Susser, 1995). Again the linearity may favour the more evidence-based areas and so limit the ability to capture learning from more innovative approaches.

In summary then, our experience with the demonstration projects would concur that, through the use of a ToC approach, planning can be sharpened and evaluation facilitated; we are less convinced about the extent to which attribution problems can be reduced. Given the paucity of literature on the practical application of theory-based approaches generally (Donaldson and Gooler, 2003), there are many remaining questions about the utilization of a ToC. First, it is important to ask whether a ToC differs from any other approach that can improve planning, such as performance management. Following on from this, should the role of improved planning and delivery be the responsibility of an evaluator? In addition, given the scale of the evaluation task implicit in the approach, can a ToC really get to the nub of *which aspects* of a complex programme of activities work for *which subgroups* of the population and in *what circumstances*? In addition, the evaluations described take place in the context of single-site initiatives at a time when in the UK and elsewhere there is a proliferation of multi-site area-based initiatives. Given the trade-off between detailed programme information and knowledge across sites, there is a question about the most appropriate evaluation context for a ToC approach to deliver its initial promise. Finally, it seems important to ask whether the investment required by a ToC produces correspondingly more useful knowledge than other approaches to complex programme and policy evaluation. These questions will only be answered through careful consideration of the evidence generated in the future by the application of the approach across varied settings and policy contexts.

Note

1. We are grateful to Professor Murray Saunders for his suggestion of the term 'the sociology of attribution' to capture this idea.

References

- Accounts Commission for Scotland (1998) *The Measures of Success: Developing a Balanced Scorecard*. Edinburgh: Audit Scotland. <http://www.audit-scotland.gov.uk>.
- Accounts Commission for Scotland (2000) *The Map to Success: Using Process Mapping to Improve Performance*. Edinburgh: Audit Scotland. <http://www.audit-scotland.gov.uk>.
- Barnes, M., E. Matka and H. Sullivan (2003) 'Evidence, Understanding and Complexity', *Evaluation* 9(3): 265–84.
- Blalock, A. B. (1999) 'Evaluation Research and the Performance Management Movement: From Estrangement to Useful Integration', *Evaluation* 5(2): 117–49.
- Brown, P. (1998) 'Shaping the Evaluator's Role in a Theory of Change Evaluation', in A. Fulbright-Anderson, A. C. Kubisch and J. P. Connell (eds) *New Approaches to Evaluating Community Initiatives: Theory, Measurement and Analysis*, pp. 101–11. Washington, DC: Aspen Institute.
- Cole, G. E. (1999) 'Advancing the Development and Application of Theory-based Evaluation in the Practice of Public Health', *American Journal of Evaluation* 20(3): 453–70.
- Connell, J. P. and A. C. Kubisch (1998) 'Applying a Theory of Change Approach to the Evaluation of Comprehensive Community Initiatives: Progress, Prospects and Problems', in A. Fulbright-Anderson, A. C. Kubisch and J. P. Connell (eds) *New*

- Approaches to Evaluating Community Initiatives: Theory, Measurement and Analysis*, pp. 15–44. Washington, DC: Aspen Institute.
- Donaldson, S. I. and L. E. Gooler (2003) 'Theory-Driven Evaluation in Action: Lessons from a \$20 million Statewide Work and Health Initiative', *Evaluation and Program Planning* 26(4): 355–66.
- Dunn, A. L., B. H. Marcus, J. B. Kampert, M. E. Garcia, H. W. Khol and S. N. Blair (1999) 'Comparison of Lifestyle and Structured Interventions to Increase Physical Activity and Cardiorespiratory Fitness', *Journal of the American Medical Association* 281(4): 327.
- Fulbright-Anderson, A., A. C. Kubisch and J. P. Connell, eds (1998) *New Approaches to Evaluating Community Initiatives: Theory, Measurement and Analysis*. Washington, DC: Aspen Institute.
- Herbert, S. and A. Anderson (1998) 'Applying the ToC Approach to Two National, Multi-site Comprehensive Community Initiatives', in A. Fulbright-Anderson, A. C. Kubisch and J. P. Connell (eds) *New Approaches to Evaluating Community Initiatives: Theory, Measurement and Analysis*, pp. 123–48. Washington, DC: Aspen Institute.
- Kaplan, R. S. and D. P. Norton (1996) *The Balanced Scorecard*. Boston, MA: Harvard Business School Press.
- Lewis, J. (2001) 'Reflections on Evaluation in Practice', *Evaluation* 7(3): 387–94.
- Pawson, R. and N. Tilley (1997) *Realistic Evaluation*. London: SAGE.
- Rush, B. and A. Ogborne (1991) 'Program Logic Models: Expanding their Role and Structure for Program Planning and Evaluation', *Canadian Journal of Program Evaluation* 6(2): 95–106.
- Sanderson, I. (2000) 'Evaluation in Complex Policy Systems', *Evaluation* 6(4): 433–54.
- Shaw, A., A. McMunn and J. Field, eds (2000) *The Scottish Health Survey*, carried out on behalf of the Scottish Executive Dept of Health. London: Joint Health Surveys Unit, at the National Centre for Social Research and Dept of Epidemiology and Public Health.
- Sullivan, H., M. Barnes and E. Matka (2002) 'Building Collaborative Capacity through Theory of Change: Early Lessons from the Evaluation of Health Action Zones in England', *Evaluation* 8(2): 205–66.
- Susser, M. (1995) 'Editorial: The Tribulations of Trials. Interventions in Communities', *American Journal of Public Health* 85: 156–8.
- Walker, R. (2001) 'Great Expectations: Can Social Science Evaluate New Labour's Policies?', *Evaluation* 7(3): 305–30.
- Weitzman, B. C., D. Silver and K.-N. Dillman (2002) 'Integrating a Comparison Group Design into a Theory of Change Evaluation: The Case of the Urban Health Initiative', *American Journal of Evaluation* 23(4): 371–85.
- Wimbush, E. and J. Watson (2000) 'An Evaluation Framework for Health Promotion: Theory, Quality and Effectiveness', *Evaluation* 6(3): 301–21.

MHAIRI MACKENZIE is Lecturer in the Health Promotion Policy Unit,
Public Health and Health Policy, Division of Community Based Sciences,
University of Glasgow, 1 Lilybank Gardens, Glasgow, G12 8RZ, UK.
[email: m.mackenzie@clinmed.gla.ac.uk]

AVRIL BLAMEY is Research Fellow in the Health Promotion Policy Unit,
Public Health and Health Policy, Division of Community Based Sciences,
University of Glasgow, 1 Lilybank Gardens, Glasgow, G12 8RZ, UK.
[email: a.blamey@clinmed.gla.ac.uk]