Aiming for Utility in 'Systems-based Evaluation': A Research-based Framework for Practitioners

John T. Grove

Abstract System dynamics modelling (SDM) was used and process researched as a case to investigate its utility as a systems-based evaluation (SBE) approach. A system dynamics (SD) model[1] was developed to evaluate the potential requirements and implications on the health systems of the ambitious antiretroviral therapy (ART) scale-up strategy in Lusaka, Zambia. Research on SDM for strategic evaluation provided insights and principles for future application of SBE. The SD diagrams readily inspired new insights while practical constraints limited use of the model for action planning. Research suggests that utility of SBE begins with engaging stakeholders to share and align their views on a representation of the system and progresses to their reinterpretations of the system that they inhabit, ultimately moving towards transformative change. Evaluators must balance two purposes in managing for utility of SBE approaches: producing a defensible representation of the system(s) and facilitating transformative change appropriately with and for system stakeholders.

1 Introduction

The publication of this *IDS Bulletin* is evidence that there is a growing philosophy, debate and a quest for examples showing that the application of systems thinking and related approaches could provide improved processes, tools and insights for evaluations and their managers and clients. However, the utility of SBE approaches has not been rigorously studied through case examples for the purpose of providing practical recommendations. SBE is the application of specific methods that are influenced by systems concepts and methodologies to the task of evaluation. This article proposes a set of recommendations for practice for evaluators and managers aiming to employ techniques derived from, or influenced by, a systems perspective.

SDM was used and process researched from 2006 to 2008 as a case to investigate its utility as an SBE approach. An SD model was developed to *ex ante* evaluate the potential requirements and implications on the health systems of the ambitious ART scale-up strategy in Lusaka, Zambia that was being developed and implemented during that period. SD was chosen, with input from the Ministry of Health, as it had been used previously for public health and disease control programme planning and evaluation and provided an accessible literature and a cadre of professional consulting resources (Homer and St Clair 1991; Homer *et al*. 2004; Homer and Hirsch 2006; Homer *et al*. 2010; Ritchie-Dunham and Mendez-Galvan 1999; Sterman 2006; Tebbens *et al*. 2005; Thompson and Tebbens 2010).

This article lays out the research stance, process and description of the case. Key features of the SD model and process are presented and described. Exemplary quotations from the research conversations before and after the SD model development among stakeholders are shared. An operational definition of SBE is proposed as well as a framework for utility as emerged from the research. Finally, implications of the research are discussed and conclusions for broader practice of SBE are proposed.

This article is based on a dissertation project cited herein and available at http://researchbank.rmit.edu.au/view/rmit:160050.

2 Research objectives, stance and process

The broad objectives of the research were to: (1) investigate the utility of SDM as to how it does or does not add value for evaluation of the ART strategy in Zambia; and (2) identify useful principles and methods for the broader application of SBE. The research questions and utility framework were informed by relevant literature, in-depth interviews, extensive field notes, and a multi-year personal research journal. Data collection and analysis for the research was guided by narrative-based inquiry and critical-hermeneutics philosophy. Specific concepts used for interpretation of data included Gadamer's (1998) role of language, Habermas's (1976) theory of communicative action, and Ricoeur's (1985) threefold mimesis. The literature review exposed key features of SBE approaches. The review suggests and reaffirms three unique core concepts in application together as essential for SBE: interrelationships, perspectives and boundaries (Midgley 2000: 45; Williams and Hummelbrunner 2010: 3).

3 SBE case study: SD model on Zambian ART scale-up strategy and context

The Zambian Ministry of Health devised a national ART scale-up strategy for 2005–9, which prioritised scale-up of services in urban sites in a first phase, with an eventual goal of decentralising to provide services at 'lower' levels of the health-care system – to district hospitals and eventually most clinics in the country (MOH 2006). Another goal was to achieve dramatic change in the epidemiological profile of HIV/AIDS in the country. Zambia's national *Scale-up Plan for HIV Care and Antiretroviral Therapy Services* strategy document established an ambitious plan to 'prevent, halt and begin to reverse the spread and impact of the HIV and AIDS by 2015' (MOH 2006: 7).

Framed as an *ex ante* strategic evaluation in support of the national ART coordinator, the SD application explored the clinical and supportive services required to enrol and sustain individuals on ART in Lusaka amid new infections at the population level and a serious lack of health-care and other ancillary workers. One approach could have been to do a goal-based evaluation – assess progress on numbers newly enrolled and sustained on ART against national targets. However, the context is more complex than that with limitations on staff and facility capacity, as

well as a clear connection between testing for presence of infection and the series of services required should one be HIV-positive. In addition, the progression of the disease is well studied and can provide a framework for gauging clinical and human resource capacity. A more nuanced definition of strategic evaluation is for: 'evaluating relationships among elements of strategy', including 'coherence of different strategy elements, complexity of inter-relationships, understanding of interdependence and inter-relationships, and evidence of systems thinking and complexity understandings' (Patrizi: 2010: 99). SDM offered a tool with a rich history, set of references and practitioners to be employed in an evaluative mode on this particular problem.

If the complexity of the Zambian ART scale-up plan and context necessitates the need for an evaluation and analytical approach that tries to encompass the dynamic context, would SD as an SBE provide greater insights and add value? Further, what is missing from our current way of doing monitoring and evaluation (M&E) of ART programmes? As such, the work was staged to include: (1) a 'baseline' round of interviews to identify the shortcomings of current M&E, early conceptualisations of the ART care system and appetite for a different approach; (2) setting priority areas and boundaries of inquiry with the national ART coordinator, conceptual and data model development, testing and production of outputs and implications; and (3) an 'endline' round of interviews which was then conducted after a series of dissemination meetings.

3.1 Baseline and preparation

Prior to model development, a round of interviews was conducted with key stakeholders to help set boundaries for the work and serve as a baseline for the utility research. Data from this phase revealed three common themes: (1) agreed critique of current M&E practice for ART and a curiosity about the potential for improving that practice through systems-oriented approaches; (2) the essential role of boundary setting for the system of interest, in terms of personal perceptions and the need to re-interpret boundaries together with others; and (3) the openness to the application of systems concepts and tools to evaluation but with a caution on new or seemingly complicated methods (Grove 2012: 160–1).

This quote, from a representative of a lead non-governmental organisation (NGO) partner supporting the national scale-up, is representative of other assertions from NGO partners from this preparatory round of interviews:

So we're good at getting the clinical stuff, and we're good at counting things and following patients, but we're bad at the more psycho/social stuff… 'Psycho/ social' is a term of ART, which, I don't really know what it means exactly. When I say it I'm talking about taking care of the patient's non-medical needs, so: counselling and making sure that the systems and structures are in place in the family, in order to take care of the family unit. From the point of view of things we measure, we'd like to know how well those other services are doing, because they affect the clinical outcomes that we're trying to deliver (Representative from lead NGO partner).

An M&E adviser from a technical partner agency notes the varieties of data collection required for a meaningful analysis of different aspects of the care system, anticipating the potential of a systems-based approach (Grove 2012: 164):

How many people have been, for example, transformed [they went quickly from gravely ill to productive once again] and have been bedridden, at time of enrolment [and have now been] transformed and working [again] and are able to support themselves? Of those that are working, we know that maybe after four or five years [those] that start having opportunistic infections, their stats start going down, how many of them have been treated? So that is one of the things that I really want to see because HIV/AIDS is no longer [solely] a health problem but it's become an economic problem. Evaluating this we would be able to answer the questions to say: this is why I'm not working and able to support myself.

Implications of this round of research suggested that proposed emphases, whether it is an emphasis that is representative of systems and related problems, or whether it is an emphasis on facilitation of transformative change, would be an ingredient to implementing an SBE. From the literature and these research conversations, SBE approaches, by design, have the potential to fulfil a need to re-interpret boundaries, perspectives and interrelationships about the ART care-and-support system and scale-up strategy in Zambia. But although this potential is there, how does an evaluation manager engage efficiently with this powerful new set of concepts and tools? The strategic evaluation of the ART scale-up plan in Zambia provided an opportunity to test an SBE approach while research was conducted alongside the process with the dual purpose of gathering formative inputs on stakeholders' conceptions of the system and to assess the utility of the potential innovation as it unfolded.

3.2 SD model scope, structure and findings

The strategic evaluation focused on the dynamics of the scale-up strategy of essential and expanding HIV/AIDS treatment services amid a raging HIV/AIDS pandemic and health-care human-resources crisis in 2006–7. Zambia faces tremendous social, public health and evaluation-related challenges. According to the Zambian Central Statistical Office's (CSO) 2004 Living Conditions and Monitoring Survey report (LCMS), 68 per cent of the total population of 10.9 million people live below the local poverty line of US$0.93 per day, with 53 per cent deemed extremely poor and therefore unable to meet the cost of basic food and non-food items. For the period 2005–9, the Zambian government devised a national ART scale-up strategy, which prioritised urban sites in a first phase, with an eventual goal of decentralising this system to district hospitals and eventually most clinics in the country (MOH 2004). Another goal was to achieve dramatic change in the epidemiological profile of HIV/AIDS to 'prevent, halt and begin to reverse the spread and impact of HIV and AIDS by 2015' (MOH 2006: 7). The scale-up plan emphasised achieving aggressive targets for new patients on therapy and identified a number of health system factors that would require significant bolstering. However, no single analytical approach had pulled both of these considerations together. As such, a dynamic and systems-oriented planning and evaluation approach had potential to provide visibility to the potentially unanticipated consequences and to potentially catalytic activities.

After consultation with the national ART coordinator and NGO partners, the following objectives of the modelling work were established:
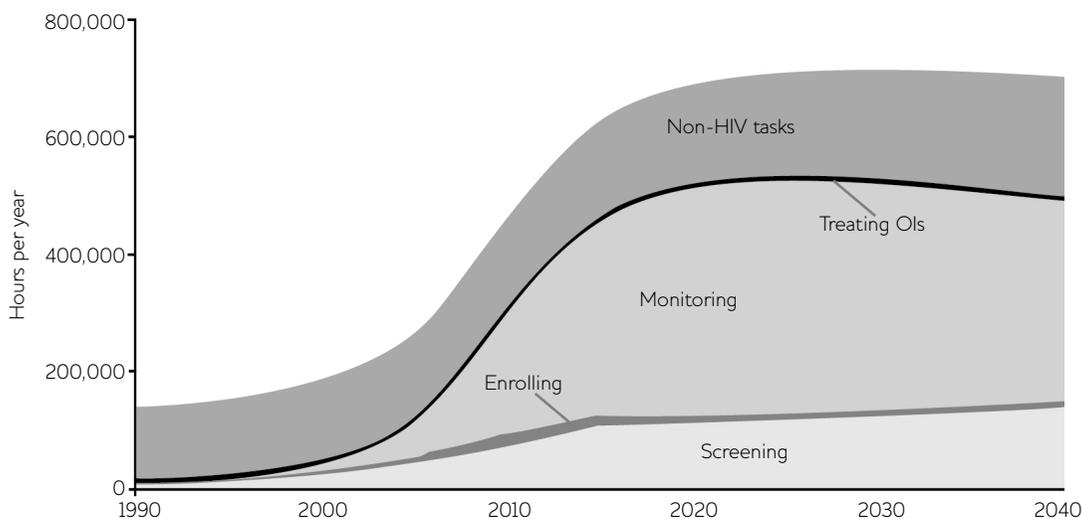
1 To assess the typical point at which the ART patient load in any given clinic has reached beyond its capacity to provide sustainable, long-term care;

Figure 1 Model overview: the 'Spaghetti Diagram'
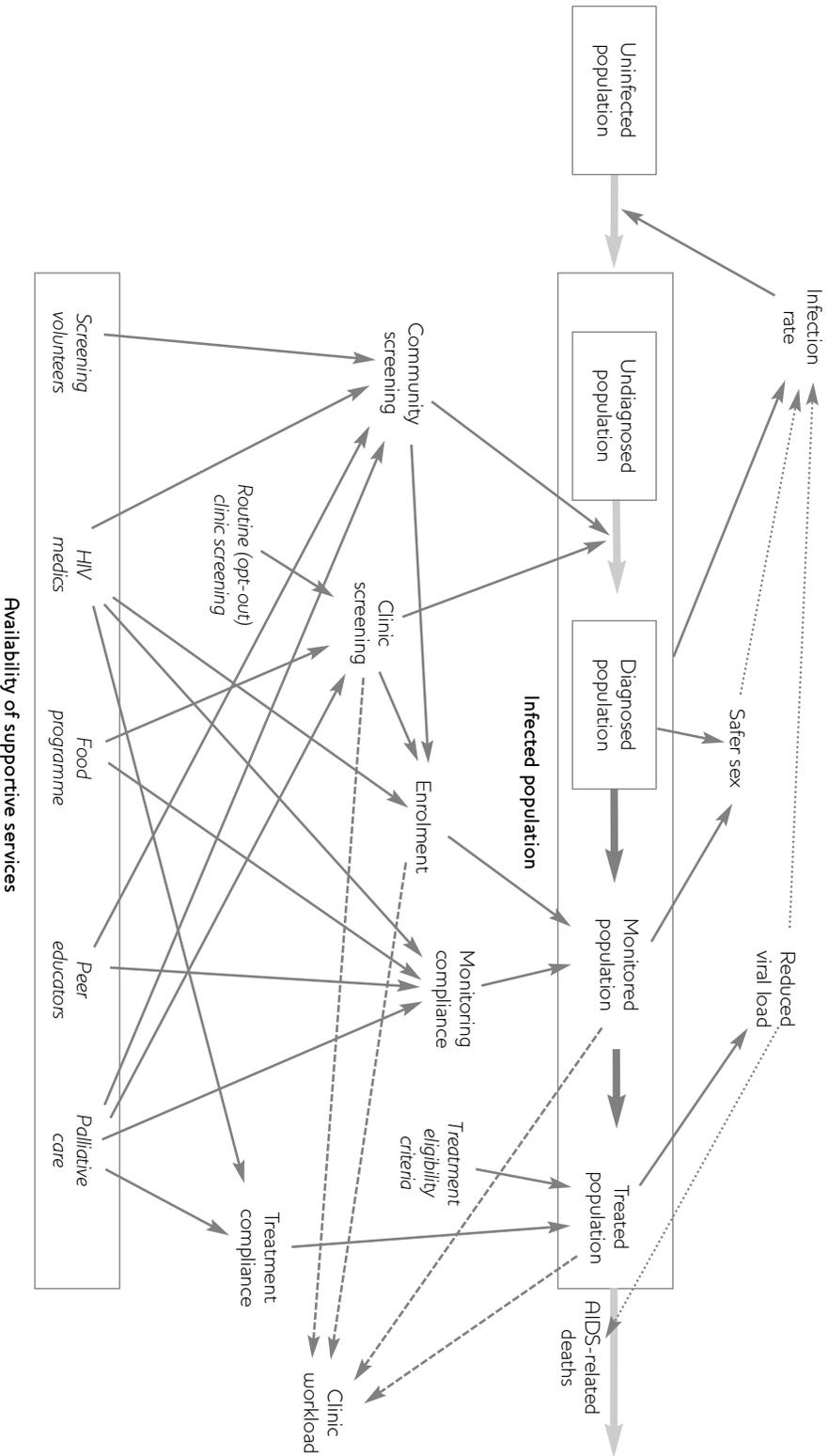
Figure 2 **The monitoring 'bulge'**

2   To determine how increasing the number and/or type of supportive services networked with an ART clinic reduces the patient load of the clinical site; and

3   To quantify what level of access to and participation in supportive health services is necessary to sustain ART for clients into the future.

3.3 *Model overview*

The model overview in Figure 1 represents the overarching structure for the model, based on a core 'stock-and-flow' diagram, moving from 'uninfected population' to various stages of HIV disease (or non-disease); a number of factors such as treatment compliance and enrolment; and the five categories of supportive services: screening volunteers, HIV medics, food programme, peer educators, and palliative-care workers serve as valves which either exacerbate or limit the number of people moving on to and staying on ART. One research participant named this figure the 'Spaghetti Diagram'.

The overview illustrates the flow of people from uninfected status, to infected status and then through various stages. The extent to which undiagnosed individuals move to diagnosis and on to pre-ART and treatment is influenced by the coverage and effectiveness of the supportive service programmes and the behaviours they foster (increased screening for HIV in the

community, clinical screening, enrolment on pre-ART, and monitoring compliance while on antiretrovirals, or ARVs). Supportive service programmes can either cause a decrease or increase in clinical workload depending on volume and tasks allocated. At the population level, the high number of treated individuals reduces viral load and drives down the number of new infections, assuming a high level of safe sex. For example, if someone is receiving food through a food programme, this is a possible outlet to promote HIV testing (clinic screening) and to promote visits for ongoing monitoring of a person's viral load, through a lab-based test of CD4 count to determine when ART should commence (enrolment and monitoring compliance in the model overview). The CD4 count is a measure of immune response cells and these are depleted when one has HIV infection. The national guidelines at the time required patients to have a CD4 count at or below 200 and/or a more grave subjective clinical staging to be enrolled in and receive ART. A portion of those screened clinically to be HIV-positive will be enrolled in ART and together with those being monitored would both add to the number of people enrolled and monitored while on treatment overall – requiring ongoing CD4 monitoring and other support services. In more recent years, the CD4 thresholds have increased to commence treatment sooner, based on evidence of effectiveness in increasing survival and reducing risk of infecting others.

This diagram served as the primary view into the causal structure for stakeholders, as well as a primary conversation piece in presentations and research conversations. A number of additional diagrams were created by the modeller, with consultation from the ART coordinator and members of the in-country technical steering group, to establish the logic of underlying areas of the system's behaviour, such as two kinds of populations – diagnosed and undiagnosed – and types of screening. Variables were determined for all of the model inputs from programme data, demographic and surveillance sources, and expert judgement on clinical-workload calculations.

3.4 Key findings

A simulation model was devised in Vensim based on the model overview structure and relevant data. Figure 2 presents perhaps the most compelling output of the SD model – the sheer number of monitored and treated patients grows over time and dramatically until late in the 2020s, because of the growth in numbers of patients who need to be treated and due to the delays in the flow from infection to diagnosis through progression of the disease to AIDS. It is important to also note that the efficacy of antiretroviral drugs in significantly reducing patients' infectivity and the ability to create new infections is a key assumption in the model. The huge number of patients that require CD4 monitoring (by laboratory test) over time exponentially increases clinician time required to provide services in terms of hours per year as well as the demand on and role of supportive service workers for the period modelled to 2040.

The place where care tasks are promoted drives people to the clinics for CD4 monitoring. At the time, national guidelines suggested a monthly visit and commencement on ART at a CD4 count of 200. In the model, supportive service workers were so successful in promoting HIV screening and compliance in the community that the clinics become quickly overloaded for enrolment and ongoing compliance monitoring. If key tasks such as HIV screening and CD4 monitoring could be decentralised, the amount of required clinician time would reduce. One pending decision during the time of the strategic evaluation was to halt efforts to start an 'HIV medics' training programme and cadre of community health providers. These are nurses or other public health professionals with additional paramedical training. However, their role was to promote testing, link community members to services, identify the potentially ill, and encourage treatment compliance. However, given that they were not able to do CD4 monitoring in the community, the patient load seeking services at the clinic is actually exacerbated as people queue for lab tests. It also shows that the number of hours spent on non-HIV tasks does not reduce but rather also increases over time as it is assumed that other illness and the burden of disease remains. Unless the number of clinicians in Lusaka were to triple or even quadruple, availability of other services would also be at risk. A goal-based evaluation may have provided an analysis of whether targets for increasing people on treatment were met and perhaps even if enough clinicians were at the treatment sites. In contrast, the SD model allowed for a potentially more insightful representation of the system, with the inclusion of factors such as task allocation, epidemiological profile and disease progression.

3.5 Strategic evaluation conclusions

Two conclusions resulted from the strategic evaluation. First, Zambia's strategic goal to 'prevent, halt and begin to reverse the spread and impact of HIV and AIDS by 2015' might be plausible for Lusaka if the huge need for additional clinical staff is met. Secondly, unless supportive service workers can perform particular tasks (e.g. screening and monitoring), their presence may actually increase patient load resulting in a potential fourfold increase in clinical workers required (Grove 2012: 214–15). One immediate decision was influenced: after presentation of the model to stakeholders, the effort to launch an HIV medics programme was de-prioritised.

3.6 Endline: model dissemination and research

The model and implications were presented to the technical steering group and the HIV/AIDS donor groups in September 2008. Research conversations were conducted with key members of these groups following dissemination workshops. In summary:

1 The model and process readily triggered stakeholders to reflect on their existing boundaries (i.e. their beliefs and values);

2 The illustrations stimulated reinterpretation of interrelationships of system components;

3   Discussion, using the overview diagram as a reference, generated concepts of new possible boundaries and components of the ART care system (such as shifting of tasks, screening and monitoring options in the community);

4   Stakeholders were comfortable with and had an understanding of the model's structure but expressed concerns with the perceived complicatedness of the approach;

5   Stakeholders expressed concern about the potential for false assumptions to be reified in the model calculations; however, they placed confidence and trust in investigators to be rigorous and that the model was as robust as possible.

A primary constraint was inability to engage with stakeholders over time on design and interpretation of the model outputs or to engage decision-makers in virtual scenario testing and refinement in a 'hands-on' way. Due to the sheer volume of work in the field required of stakeholders, that is, building the ART system and providing services, engagement options were limited to rounds of feedback on prototype representations of the system and sharing of model outputs. Facilitation of further use of the model for decision-making was severely delayed and ultimately cancelled due to a financial crisis and staff turnover at the Ministry of Health. More intensive stakeholder engagement in conceptual design both *during* the development and *after* the completion of the model would have been optimal. However, the reality of the ART scale-up placed tremendous demand on people's time. So the trade-off was to develop drafts for review and consultation, as opposed to a group model-building approach.

Overall, the introduction of the SD model structure and results readily triggered stakeholders to reflect on their existing boundaries; specifically their beliefs and values about the role of various support service workers and the level of ambition of the scale-up. The model also stimulated reinterpretation of interrelationships between service cadres, including new options for the ART care system. One Zambian official noted:

My general observation is that it is a good model. I think when I got back to the office and I was, sort of,

trying to go through [the presentation hand-outs]. And when I see these materials [print-out of presentation slides], I think one thing that I observed is that it is actually very important in the sense that it, sort of, highlights issues that maybe we don't [usually] directly link to the ART services. It's sort of showing us the other side of the coin on issues that actually affect [the] provision of ART services, which we don't look at. And even when we do, for instance, [look at] the evaluations and even the annual reviews, we don't actually focus on those issues to see as to how are they actually driving the whole programme. So that was actually the thing that came out. It was very, very important, and I think it's useful. It's an eye-opener in terms of when you're doing evaluations for these programmes, what should we focus on as opposed to the 'business as usual'.

This was a typical reflection on the overview and outputs of the model. Additional data suggests that stakeholders had comfort with and an understanding of the model's structure, but they also talked about concerns with the number of assumptions used and a desire to get more engaged in the details of the work so as to 'trust' the model outputs.

You can't just assume that what it comes [out] with is the truth. You have to keep on going back and checking what is put into it again before you [can] say, 'Okay, fine, this is right', because, as we said, when you look at [the year] 2008, you realise, 'No, no, no; this is way too high, or way too low'. You have to go back and try and peek [i.e. look at initial simulations] before you actually get something out [which can be trusted].

While expressing concern about the potential for false assumptions to be reified in the model calculations, participants, at the same time, placed confidence and trust in the investigators as long as limitations were communicated in written and verbal presentation.

I guess, looking at the whole process and the rationale behind the whole model, I think – right now I think it's acceptable, but in the long run, as the model is refined, I think that's when one really needs to go into the nitty-gritty. I guess now, […] maybe asking the people who see the way it operates, and maybe you as a pioneer – maybe the best thing right now is to really understand the nitty-gritty, because it's a model. So as we use it, then it will sort of unfold a number of issues that we really have to consider. So to that effect [to be

able to consider scenarios], I think right now that would be acceptable, but I think the use of projections and assumptions is allowed. The good thing is that whatever the assumption that you are using, it's also supported with some programme data [from the Zambia ART programme]. So, I think for that level, I think it's acceptable. But in the long run, you would need to zero in on the specifics.

The model provided a useful representation of the ART care system and stimulated learning on the scale-up strategy; however, more transformative or sustained changes to programme approaches did not occur immediately. Further use of the model would have possibly allowed individual stakeholders to: (1) assert their claims on validity of opinions, data and proposed action towards a shared 'is'; (2) re-interpret boundaries, interrelationships and perspectives towards a shared 'ought'; and (3) move towards consensual action for system change. This is consistent with the research and documented practice on SD (Barton and Haslett 2007; Homer 1996; Homer *et al.* 2004; Homer and Hirsch 2006; Huz *et al.* 1997).

There was disagreement among participants about in-country capacity to do 'home grown' SD modelling and similar SBE in Zambia. Some felt strongly that with minimal training and institutional ownership this capacity could easily grow, while others felt that the model's level of abstraction and initial perceived complicatedness would necessitate significant time, effort and resources to build local capacity in SD (Grove 2012: 246–96). This is a potential constraint yet opportunity for evaluation capacity-building; however, the evidence base of useful SBE applications would strengthen their validity and provide examples for learning and adaptation.

Key themes of the Zambia SDM case suggest broader implications for practitioners employing SBE. For example, participants asked for more time and opportunities to understand modelling and analysis. Capacity should be built with adequate time to train participants in new language and concepts. Greater understanding of how data are manipulated and more opportunity to engage with simulations matched to burning questions must also be facilitated. There is also a need to collect a broad variety of perspectives at the very outset of the concept development and modelling process.

Process aims should be determined and communicated at the outset. Is the aim to develop a rich description of the system and consider new components or operational bottlenecks and stop there? Or is the aim to develop that rich description and utilise it for action planning and transformation of the system? SBE demands that the planner/evaluator or manager emphasises explicit representative and/or transformative process aims at the outset and plan technical and 'softer' pieces of the overall process with detail.

Determining the system's boundaries is a *methodological asset* for, and of, many SBE approaches. The SD model offered a point of reference for dialogue which enabled the generation of new and unfolding understandings through dialogue with potential to lead to *coordinated action*. 'This (dialogue) can serve as an opening to guide groups of stakeholders who are oriented towards reaching understanding with each other to share their claims about truth and solutions for coordinated action' (Habermas 1976).

4 Utility re-visited: lessons suggest broader implications

SBE is the application of specific methods that are influenced by systems concepts and methodologies to the task of evaluation. Borrowing from action theory, these methods articulate and analyse the *interrelationships and interdependencies of real-life factors*, including perspectives and boundaries, in a process of intervention towards action (Grove, Kibel and Haas 2006; Midgley 2000; Imam, LaGoy and Williams 2007). More specifically, the broader literature and this case study research suggest that SBE methods have potential to:

1 **trigger** stakeholders to make explicit, known and clear their pre-understandings about problems and their placement and interrelationships in the system (one's 'is');

2 enable, through language, **comprehension of issues about the broader system (everyone's individual 'is')**;

3 promote or engage stakeholders' assertion of their **claims on validity of opinions, data, and proposed actions (getting to a shared 'is')**;

4 catalyse **expanded reinterpretation** of relationships, interrelationships and

boundaries (getting to a shared 'ought') on an agreed way forward;

5 optimally facilitated, move stakeholders to consensual **coordinated action (what/when/how)**.

Put simply, for effective SBE the practitioner facilitates a set of engagements with stakeholders across a continuum: Trigger, Comprehend, Validate, Re-interpret, and Act. The work is phased to align pre-understandings, developing a shared comprehension of the dynamic situation and components, and providing space for argument on the validity of opinions and data. This is consistent with Hummelbrunner's article in this *IDS Bulletin* in the promotion of moving from single- to double-loop learning, where SBE has potential to expand the repertoire of approaches to move from single- to double-loop learning – and potentially third-loop learning if facilitated with power dynamics at the core of investigation.

If alignment and reinterpretation are realised, action can be planned. SBE can be facilitated with an aim to move stakeholders to consensual *coordinated action* (Habermas 1976; Midgley 2000).

Based on the case study research and features outlined above from other examples, a simple framework for considering the utility and related trade-offs of employing SBE emerges. Utility also lies in: (a) how the representation of a system is developed; and (b) the extent to which deliberate facilitation towards transformative action is carried forward with that representation in mind. Figure 3 illustrates the dual notions of utility for SBE – the representative role of a systems approach nested within the process of transformation as an overarching aim along the Trigger, Comprehend, Validate, Re-interpret, Act continuum. With an intention to be 'representative', the practitioner aims to develop suitable observations and investigations to produce a description of a system and test them. The practitioner can then use that representation to develop a transformative process by facilitating in-depth dialogue for understanding and improving the system of, by, and for system stakeholders.

To elaborate, the distinction can be understood through these two contrasting hypothetical statements:

Figure 3 **Overarching elements of utility**

Transformative
Facilitated in-depth dialogue for understanding and improving the system of, by and for system stakeholders

Representative
Observation and investigation to produce a description of a system
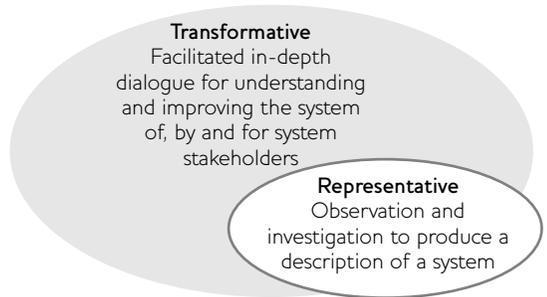
Source Author's own.

1 Now that I understand that the problem is a system, I can see or consider (evaluate) what's needed to possibly change it. I may be the one to make the change or I can help advocate for someone else to do so. (Representative)

2 I was part of describing the problem and the system, I am part of the system, I care about the system, and I value certain factors and connections. I will now go create and be part of the new system. (Transformative)

The evaluator or manager must then establish process, and gain or contract expertise to mediate these two intentions on a spectrum from aligning pre-understandings among system owners, developing shared comprehension of the dynamic situation and components, providing space for argument on the validity of opinions and data and reinterpretation of interrelationships, perspectives, and boundaries, and planning for action.

5 Implications for SBE practice
Emphases for launching and guiding an SBE approach by an evaluator or evaluation manager can be guided by the clarification of the desired 'process outcome' and its intended uses as guided by the two modes of practice: (1) representative and/or (2) transformative. In most cases, both modes are being pursued such as developing the representation to aid in transformation. Considering the features of each mode can help to clarify where to begin and how to proceed.

For each mode, there will be (a) identified situational needs driving methodological choices, (b/c) differing emphases of our ontological and epistemological roles as evaluators, (d) a spectrum of scope-setting factors related to boundary

Towards desired process outcomes	Representative Observation and review of evidence to produce a shared description of a system.	Transformative Facilitated, in-depth dialogue for understanding and improving the system of, by, and for the system's stakeholders.
(a) Situational needs	Representation of system created from dialogue, observation and evidence.	Co-created meaning of system and its needs through intensive dialogue.
(b) Ontological role	Categorise the nature of the system, based on evidence.	Explore to understand the nature of the system, as understood and conveyed among stakeholders.
(c) Epistemological role	Compile evidence and observations to create a representation of a system(s).	Compile evidence from stakeholders to generate revised or updated observations through iterative dialogue.
(d) Boundary critique	Use existing frameworks to set pre-understood and generative boundaries.	Explore privately held, and generate new, boundaries in dialogue with individuals and groups.
(e) Evaluation design factors	Establish specific questions and gather observations and/or evidence.	In-depth consultation with wide stakeholder input into design.
(f) Strengths/limits Application of method	Time-bound and efficient. Potential exclusion of stakeholders. Resource intensiveness can vary.	Broad stakeholder engagement. Time-consuming; resource-intensive process.
(g) Capacities required for implementation	Technique proficiency. Security of productive working relationship(s). Agreement on process/direction.	Group facilitation techniques. Continuous management of broad and numerous stakeholder relationships.

Source Author's own.

critique and boundary setting, (e) overarching evaluation design and implementation factors, (f) specific tendencies for application of methods and related strengths and limitations, and (g) required and preferred capacities necessary for implementation, whether by an individual or a team. Table 1 sets out descriptors that point to key elements of these two modes of practice. Situational needs, the evaluator's ontological/epistemological roles, boundary critique, design factors, strengths and limitations, and required capacities are outlined for each.

5.1 Representative mode
The representative mode emphasises the necessity of purposive observation and a review of evidence to describe or depict the system. This mode is emphasised in a situation where the stakeholders aim to understand the problem, system, and programme dynamically through gathered evidence/observations, in order to be able to represent the system's interconnected problem drivers and issues. The evaluator's

ontological role focuses on questioning and on the categorisation of evidence on the nature of the existing system and proposed programme. These categories are organised in a way that enables reflection among stakeholders on current and alternate, future realities.

5.2 Transformative mode
The transformative mode emphasises the necessity for meaning-making for understanding and improving the system by the system owners, based on the defensible representation of the system. The process becomes an intervention of its own in the group or community. This mode serves the evaluation aims to understand the problem, system and strategy or programme through inquiry into possible actions on system factors. The evaluator's ontological role is to explore issues and the imagined effects of intervention programmes within systems, in order to understand the shared nature and experience of reality as it is understood among stakeholders. The evaluator's epistemological

task is to explore the observed effects of programmes operating within systems, in order to compile evidence from stakeholders that is relevant to the issue, such as the ART scale-up strategy, and programmatic interventions at play. The evaluator then uses the observations to determine action steps for change.

A key strength of this mode is the broad and in-depth stakeholder engagement throughout the process, including its design, process and planning for action based on its results. This mode makes it clear that those who are engaged are, indeed, themselves, the very change to the system. The limitation of this mode is that such engagement can be time-consuming, resource-intensive with respect to person-time and facilitation costs, and subject to situational changes and a shifting focus.

5.3 Re-considering the Zambia case

Determining how to balance representative and transformative modes of practice can guide the evaluator to an appropriate mix of SBE approaches. It is important to note that these approaches will likely have numerous attributes and any given evaluation project may cover different issue sets depending on its scope.

For example, in the Zambia case, the representative/epistemology (1c) aspects were maximised, while the (2b) transformative/ontology aspects could have been more aggressively integrated, by using the simulation in the post-model period as a tool for discussion and reconsideration of programme designs or policy changes in the future. Data illustrated that participants could engage with the 'spaghetti diagram' (2bc), but that they expressed less comfort with the model's input equations, results and figures (1bc).

In some cases where an evaluator is interested in applying a systems approach to evaluation, their point of departure for choosing a method would be obvious but perhaps combinations would be necessary to achieve the representative or transformative purpose. For example, use of Ulrich's Critical Systems Heuristics (1986) for boundary setting emphasises the transformative/ontology elements (2b) as a first step, but this could easily lead to representative/epistemology (1c) formats for longer-term data collection and consideration, once the initial understandings of

stakeholders became unified. The evaluation design phase/s (e) is important here, in order to determine the extent of the data collection and analysis that will be flexible and based on changing circumstances, versus those items that are more fixed and non-negotiable. Moreover, the way in which methods are employed (whether 2e or 1e) has a direct influence on the extent to which results are considered valid and whether the evaluation is completed efficiently. Boundary critique (d) is an explicit aim for Ulrich and a core functional aspect of SBE; however, boundary critique can be employed to emphasise (2d) the generation of new understandings with regard to beliefs and values, or to focus on (1d) existing boundaries (or, which is likely, a bit of both).

The (g) capacities required of the evaluator or evaluation team for the project (at least at different points in the project), would depend upon the emphasis of (1) representative and/or (2) transformative outcomes. An SBE which is heavy on stakeholder engagement in groups, for example, would require the evaluation team to include someone well versed in group-process techniques and their relative strengths and weaknesses (2f). Conversely, certain stages may require an emphasis on analysis that is independent of group meetings, and the skill set required would need to match the methods used, in addition to an ability to articulate the process, trade-offs and communication of the information produced. The evaluator or team has to call upon a mature sense of their ontological and epistemological stances at a personal level, in order to be able to then mediate among participants and sponsors throughout the project towards the purpose.

Considering Table 1 in reference to an SBE effort can serve as a kind of litmus test to match the potential usefulness of certain methods to the system's needs that have been determined (by stakeholders) from a *practice* perspective rather than a sanctioned relational chart with one-to-one matches of situation to method. By considering this framework, evaluators and managers can determine which emphases fit the situation they are studying, and they can proceed accordingly with exploring methods in consultation with the available literature and methodological specialists.

6 Conclusion

SBE potentially offers a new horizon for the field of evaluation. Evaluators and programme sponsors who want to pursue real-world complex issues with methods that embrace rather than aim to deconstruct complexity should continue to contribute to building practice and understanding of SBE. While admittedly this is one case, the Zambia experience illustrates that even the use of a single system diagram can trigger learning for evaluation.

Table 1 is not an attempt at a fixed taxonomy but, rather, it is an attempt at identifying key elements of SBE and at provoking further discussion in the field. As Imam *et al.* caution about the lack of taxonomies in the fields of evaluation and systems: 'a framework that joins the two fields is a bit like pinning a tail to a tail – there would be no substantial body to support it from either field' (2007: 212). The authors go on to note that it is too early for such taxonomies and say: 'let's see how the relationship develops in practice for a couple of years before beginning to codify it' (*ibid.*).

As such, more case examples are needed to build a field of practice – one that is informed by practical experience that tests the various philosophical roots at play – yet coupled with an emergent set of principles and quality elements. Continued research and communities of practice on specific applications – identifying what 'helps' or 'hinders' utility – would build a useful evidence base for evaluators, researchers, systems practitioners and programme implementers. A compendium of resources, organised by type of SBE approach would also begin to de-mystify the principles and processes associated with systems and evaluation techniques alike. Efforts to compare theories and uncover commonalities through shared philosophical roots between fields can also provide insights and ensure that integrity of method and practice is ever-strengthened to improve the lives and systems of beneficiaries.

Note

1 The SD model was created by Dr Jack Homer.

References

Barton, J. and Haslett, T. (2007) 'Analysis, Synthesis, Systems Thinking and Scientific Method: Rediscovering the Importance of Open Systems', *Systems Research and Behavioral Science* 24.2: 143–55

Gadamer, H.G. (1998) *Truth and Method*, Joel Weinsheimer and Donald Marshall, trans., New York NY: Continuum Press

Grove, J.T. (2012) 'Utility and Practice of Systems-based Evaluation: A Case Study of Health Service Expansion in Lusaka', Zambia, Dissertation, Royal Melbourne Institute of Technology, Australia, http://researchbank. rmit.edu.au/eserv/rmit:160050/Grove.pdf (accessed 13 October 2014)

Grove, J.T.; Kibel, B. and Haas, T. (2006) 'EvaluLEAD: An Open-Systems Perspective on Evaluating Leadership Development' in K. Hannum, J. Martineau and C. Reinelt (eds), *Handbook of Leadership Development Evaluation*, San Francisco CA: Jossey-Bass

Grove, J.; Homer, J.; Mwenda Mwapela, L.; Zulu, I.; Stringer, J.; Chi, H. and Mwango, A. (2010) 'Task-shifting to Whom? The Potential Role of Supportive Service Cadres for Antiretroviral Therapy (ART) Sustainability in Lusaka, Zambia', unpublished report, WHO Health Systems Research Symposium, Montreux, Switzerland, 16–19 November

Habermas, J. (1976) *Communication and the Evolution of Society*, Thomas McCarthy, trans., Boston MA: Beacon Press

Homer, J. (1996) 'Why We Iterate: Scientific Modeling in Theory and Practice', *System Dynamics Review* 12.1: 1–19

Homer, J. and Hirsch, G. (2006) 'System Dynamics Modeling for Public Health: Background and Opportunities', *American Journal of Public Health* 96.3: 452–8

Homer, J. and St Clair, C. (1991) 'A Model of HIV Transmission through Needle Sharing', *Interfaces* 21.3: 26–49

Homer, J.; Hirsch, G.; Minniti, M. and Pierson, M. (2004) 'Models for Collaboration: How System Dynamics Helped a Community Organize Cost-effective Care for Chronic Illness', *System Dynamics Review* 20.3: 199–22

Homer, J.; Milstein, B.; Wile, K.; Trogdon, J.; Huang, P; Labarthe, D. and Orenstein, D. (2010) 'Simulating and Evaluating Local Interventions to Improve Cardiovascular Health', *Preventing Chronic Disease* 7.1: A18

Huz, S.; Andersen, D.; Richardson, G. and Boothroyd, R. (1997) 'A Framework for

Evaluating Systems Thinking Interventions: An Experimental Approach to Mental Health System Change', *System Dynamics Review* 13.2: 149–69

Imam, I.; LaGoy, A. and Williams, B. (2007) 'Introduction', in B. Williams and I. Imam (eds), *Systems Concepts in Evaluation: An Expert Anthology*, Point Reyes CA: EdgePress

Midgley, G. (2000) *Systemic Intervention: Philosophy, Methodology, and Practice*, New York NY: Kluwer Academic/Plenum Publishers

MOH (2006) *Scaling-up Antiretroviral Therapy for HIV/AIDS in Zambia: National Scale-up Plan 2006–2008*, Lusaka: Ministry of Health Government of the Republic of Zambia

MOH (2004) *Scaling-up Antiretroviral Therapy for HIV/AIDS in Zambia: National Implementation Plan 2004–2005*, Lusaka: Ministry of Health Government of the Republic of Zambia

Patrizi, P. (2010) 'Strategy Evaluation: Emerging Processes and Methods', *New Directions in Evaluation* 128: 87–102

Ricoeur, P. (1985) *Time and Narrative*, Vol. 1, Kathleen Blamey and David Pellauer, trans., Chicago IL: University of Chicago Press

Ritchie-Dunham, J.L. and Mendez-Galvan, J.F. (1999) 'Evaluating Epidemic Intervention Policies with Systems Thinking: A Case Study of Dengue Fever in Mexico', *System Dynamics Review* 15: 119–38

Sterman, J.D. (2006) 'Learning from Evidence in a Complex World', *American Journal of Public Health* 96.3: 505–14

Tebbens, R.J.D.; Pallansch, M.A.; Kew, O.M.; Caceres, V.M.; Sutter, R.W. and Thompson, K.M. (2005) 'A Dynamic Model of Poliomyelitis Outbreaks: Learning from the Past to Help Inform the Future', *American Journal of Epidemiology* 162: 358–72

Thompson, K.M. and Tebbens, R.J.D. (2010) 'Eradication Versus Control for Poliomyelitis: An Economic Analysis', *The Lancet* 369.9570: 1363–71

Ulrich, W. (1986) *Critical Heuristics of Social Systems Design*, Department of Management Systems and Sciences, University of Hull Working Paper 10, Hull: University of Hull

Williams, B. and Hummelbrunner, R. (2010) *Systems Concepts in Action: A Practitioners Toolkit*, Redwood City CA: Stanford University Press

Williams, B. and Imam, I. (2007) 'Systems Concepts in Evaluation: An Expert Anthology', in M. Scriven (ed.), *The Monograph Series*, The American Evaluation Association, Point Reyes CA: EdgePress