Coping with Complexity: Decision Analysis in Public Health[1]

Kenny Lawson[1], Elisabeth Fenwick[1,2], Andrew Briggs[1] and Lyndal Bond[2]

[1]Public Health and Health Policy, University of Glasgow, Glasgow, UK
[2]MRC Social and Public Health Sciences Unit, Glasgow, UK

Abstract

While public health interventions may be more complex than health technology policy makers are still faced with choosing between uncertain alternatives in making funding decisions, given the reality of opportunity cost. Consequently, the application of Decision Analysis as an approach to prioritisation remains valid. This also provides a focus to identify key fundamental research gaps to improve both the information set and evaluation methods. The paper discusses four main areas where further research would be useful, including: further development of common outcome measures; progress in formalising causal intervention theories for modelling; improvement in the development and application of more robust study designs; and a more formal approach to defining the conditions under which findings can be generalised to other contexts. Evaluations need to uncover not only effect sizes and costs, but also how an intervention worked and for whom.

[1] Work in progress: Please do not cite without prior permission.

1. Introduction/Background

1.1 Demand for HE to get involved*:* There is renewed interest in public health (PH) interventions; with specific focus upon demonstrating "value for money" (Darzia 2006, Wanless 2004). Consequently, the role of NICE was widened to consider PH interventions (2006) and, in recognising the intersectoral nature of PH interventions, the National Institute for health research (NIHR) is interested in methods to cope with complex interventions of non-NHS interventions (2008).

1.2 Current supply of evaluations of complex interventions*:* Economic evaluations have been increasing over the past decade. Most have concentrated on downstream (health behavioural change); and have used QALYs (McDaid, 2007). There has been limited involvement in assessing the effectiveness of upstream interventions or consideration of appropriate methods to adopt in PH evaluations, where the interventions tend to be significantly more complex than in health technology (Shiell 2007, Kelly 2006). A recent review of the applicability of HTA methods to complex PH intervention identified several areas of research (Drummond et al 2006).

Non-economic approaches dominate the evaluation of complex interventions. Typically, there is a concentrate on complex description of "how" and "why" an intervention worked. There is therefore scope for health economics to add value by looking at effect sizes and costs. In particular the concept of opportunity cost is not considered; the implicit assumption is that the intervention is "good value". Typically the focus is "within programme" evaluations; motivated to improve learning-by-doing. At the extreme, empirical approaches can be are dismissed out of hand as philosophically naïve(MacPherson 2007, Pawson 2004); counterfactual and controls groups are viewed at best with scepticism, and consideration of value for money is ignored.

1.3 Outline of the paper The premise of the paper is that the approach of Decision Analysis provides an ideal framework to approach and undertake evaluations of complex PH interventions. The contention is that despite complexity we are still choosing (at least implicitly) between uncertain alternatives when making a decision to fund a PH intervention. Evaluation should therefore provide adequate information to decision makers.

This is essentially a discussion paper highlighting key research issues, rather than reporting upon a particular empirical study, or making prescriptions. In addition, given the recent attention by health

economists to public health there are significant opportunities for both methodological and theoretical development. Therefore, the pitch of the paper is to outline perceived key gaps, rather than focus on a particular area. The paper begins by briefly outlining the Decision Analytic approach; it then illustrates what is meant by complexity in public health interventions, and presents some of the implications of this complexity for evaluations. For illustration, we also reference several evaluations currently underway (or recently completed) at Glasgow. Finally, we turn our attention to identifying four main areas of research that can usefully be enhanced to improve the usefulness of evaluations for decision making. Namely, the development of wider, *and common*, outcome measures; modelling the theory of intervention and in particular the potential utility of "system approaches"; development and applications of a wider menu of study designs, and finally more explicit consideration of the limits of generalisability[2].

2. Decision analysis and complex PH interventions

2.1 Brief summary of Decision Analysis: It is worthwhile giving a brief overview of Decision Analysis (DA) given its application is the central premise of the paper. DA is the *general approach* to structure a decision problem when choosing between uncertain alternatives; where prioritisation is unavoidable and opportunity costs are inherent. DA has been applied to wide range of fields from engineering, operational research and business strategy. It is routinely used in health technology assessment, and our premise is that it is equally applicable to public health decisions. The approach itself seeks to make explicit the key elements of a decision including the maximand; causal variables and relationships; the informational set; and finally the uncertainties associated with each[3]. Therefore, a variety of methods can be used depending upon the decision problem at hand. In short, decision analysis is a modelling exercise which makes explicit an otherwise complex and implicit decision process. Crucially, this transparent approach is reproducible, open to peer-review, and can be applied on an iterative basis; allowing decisions to be updated as new and valuable information is received. "All models are wrong"[4], but Decision Analysis is intended to improve upon an otherwise implicit process, based upon mental models of heuristics and subjective experience. The value judgements are the decision makers, and Decision Analysis seeks to provide a rational framework, transparency and therefore accountability to policy choices.

[2] Drummond et al outlined four areas for research: wider outcomes, causal methods, intersectoral impacts and equity; though without goes into details as to possible future approaches. Within this paper, consideration of common outcomes is very brief, and it is hoped there is value added in the areas proffered for consideration.

[3] Decision analysis is simply an (epistemological) approach to making explicit the different elements of the decision process, and is "value free". Normative judgements are the proviso of decision makers and are essentially inputs into the maximand.

[4] George Box.

2.2 What makes PH intervention complex? PH interventions wide and varied, ranging from population immunisations and screening to regeneration to transport. The following section illustrates the main stylised elements of why certain PH interventions and the evaluations of them are termed as being "complex". PH interventions are/can be:

- Similar in name (e.g. exercise counselling), but different in nature; varying in specific design (e.g. skill of counsellor, individual versus group; intensity, duration, follow-up etc)
- Programmatic and multifaceted: multiple and simultaneous interventions;
- Involve different sectors/institutions; levels of government and non-government organisations;
- Where implementation may not reflect design;
- Differ in focus on short and long term impacts (extending to intergenerational impacts);
- Target a variety of target audiences (individuals, communities, populations);
- Evolve over time (design, delivery) through learning-by-doing and feedback ;
- Predicated on avoiding the incidence of illness. This complicates equity issues in raising tensions between individual responsibility and history (e.g. past behaviour) versus paternal outcomes;
- Differ in scale and therefore impact. Even in theory the "dose-response" relationships and mix of policies necessary to achieve the desired impact are not well articulated (e.g. issues of critical mass, tipping point time to impact etc);
- Have wider sectoral consequences (outcomes and costs), and outcomes can exist at the level of the group or community, rather than the individual (e.g. social capital);
- Are implemented in a changing context. Context is important - acting as both a confounder of impact and is an effect modifier by interacting with the causal drivers of effectiveness. Context varies in space and time thereby the generalisability of PH intervention is likely to be limited.

Context is central theme within the paper and we primarily refer to the wider determinants of health (economic, social, environment and legal) as well as to simultaneous mix of relevant upstream (e.g. policy changes) and downstream (e.g. behavioural change). This is typically referred to as "distal context" in other social disciplines. On the hand, influences on programmatic level we will refer directly as design and implementation features; where this can commonly be referred to as "proximal context". This further raises the sociological/ecological observation that

society comprises of multiple interaction "systems"; thus raising concerns that one must account for heterogeneity at the both the spatial and individual level – for instance one must avoid the "ecological and atomistic fallacies" (see Leyland et al 2003).

PH interventions are typically complex in different ways. At Glasgow, there are three main evaluations in the area of public health: SHARP, GOWELL and KeepWell[5]. The first has just been completed was a Scotland wide housing development and refurbishment programme. The second is a housing and regeneration programme across Glasgow, with multi-sites defined by different intervention mix and levels of intensity (from new homes, refurbishment, new amenities, community initiatives, increasing the accessibility to the city centre for peripheral areas etc). Finally, Keep Well is a national CVD prevention programme for both primary and secondary patients. The former are identified through screening 45-64 years in deprived communities for high risk of CVD ($\geq$ 20% over 10 years) and offering preventive behavioural lifestyle and pharmacy advice. We will be referring to these programmes later. The point in this section is that PH interventions can be complex in different ways. Appendix 1, illustrates that while housing and regeneration interventions often have a wide range of complex outcomes, the intervention design and delivery itself can be relatively straightforward (although more complex for GoWell). Conversely, behavioural change programmes such as the prevention of cardiovascular disease typically involve a huge array of behavioural and pharmacological interventions, and requires long term extrapolation of impact; however the outcome of interest (CVD) is simply defined.

2.3 Implications of complexity for PH research? The evaluation of PH interventions contrast sharply with HTA: where the epidemiology is well developed; interventions are *relatively* simple; heterogeneity more easily defined; and generalisations more straightforward. In public health, information and robust methods are quite weak. This provides a daunting challenge to evaluators. However, Decision Analysis is an approach to begin to compare relative effects (and uncertainty) between competing PH interventions[6]. The main implications of complexity for research is to improve: (a) the choice of outcome measure(s) to be used; (b) the need to articulate the theory of intervention; (c) to adopt the most robust experimental designs; and (d) generalisability: formalise the conditions under which social technology can be transferred.

[5] SHARP http://www.sphsu.mrc.ac.uk/study-sites/sharp/; GOWELL http://www.gowellonline.com/; Keep Well http://www.healthscotland.com/Prevention-2010.aspx

[6] In this paper we side-step the operational issues that budgets to fund PH interventions are not infinitely fungible between sectors and levels of government. This is a vital, but separate question.

3. Key research area I: wider and common outcome measures

To help policy makers prioritise consistently across PH interventions, we need "a common measuring stick": to create common *final* outcomes measures across PH interventions. Ultimately, which outcomes are selected must come from legitimate decision makers[7]. The researcher's role is to aid with the articulation of this process. The role of the evaluators is then to hold the intervention to account using agreed objectives. Common outcomes measures allow this process to be transparent. Projecting to common outcomes may serve as a basis to choose between innovative projects.

There is now a range of on-going work into wider and common outcomes measure including the Capability approach, and there are several on-going projects that are essentially "proof of concept" (see Coast et al 2008). An operational framework remains a medium to long term possibility. Given the need to compare interventions with impacts across short versus long term time frames; different aspects of well-being; variations in populations (etc) are we able to generate "a common measuring stick"? Eventually, alternative welfare (well-being) frameworks will come up against the same problems encountered in standard welfare economics – e.g. cardinal versus ordinal measurement; interpersonal comparisons; adaptive expectations etc.

In the short term, perhaps the use of QALYs supplemented with a cost consequence analysis is a short term fix. However, cost consequence does not help transparent prioritisation in Decision Analysis and there are a range of prioritisation techniques available to weight the relative importance of outcomes including: Multi Criteria Decision Making (MCDA); Analytical Hierarchy Process (AHP); Discrete Choice Experiment (DCE); and Best Worst Scaling (BWS). These ideas are well rehearsed; given length constraints the paper now concentrates on areas given less attention.

4. Key research area II: Further develop the theory of intervention

In contrast to health technology assessment (HTA), social epidemiology (and aetiology) is underdeveloped for PH interventions: the causal mechanisms and relationships in public (and social) interventions are often poorly articulated and even less well understood. This section outlines why a theory of intervention is important and then illustrates the emerging approach of systems theory and causal mapping.

[7] Where equity judgements typically are derived from population preferences.

4.1 Why a theory of intervention is required: Perhaps there is an implicit assumption that if there is a good study design and well specified control groups then effect sizes will be valid, and generalisable. If so, then why bother with a theory of intervention, if we know it "works", or "doesn't". It is contended that while study design is critical, without (i) a well specified theory of intervention our knowledge of counterfactuals and selection of adequate control groups may be flawed; (ii) even in a randomised design, without knowledge regarding the influence of contextual variables presumed effect sizes may be non-stationary in different contexts (time/space), and generalisations are likely to be limited; (iii) without combining process evaluation to determine implementation success one may confuse a poorly designed intervention from a poorly implemented one (Hawe, 2004). These reasons are expanded upon in the next section.

The literature is very familiar with techniques in process evaluation. The focus of this section is to provide a summary of a promising approach of System Dynamics and its potential application in developing a theory of intervention and a basis for quantitative modelling. This is intended to shed light on the "black-box" of context; as an aid to endogenously capture the variety of upstream and downstream influences that can have a direct effect on the intervention itself and serve as confounders.

4.2 Systems dynamics and causal mapping approaches:

Systems Dynamics (otherwise known as System Theory or Systems Modelling) is a general equilibrium approach to causal inference which is applied when there are multiple and simultaneous influences on outcomes of interest. The main purpose is to endogenise key variables that would otherwise be held as exogenous which could lead to spurious inference (Sterman 2000). For instance, it is not uncommon to read evaluations which adopt a study design of non-randomised controls and list a range of possible factors that may account for a null result. Systems Dynamics is an approach to incorporate these drivers into the theory and modelling of the intervention. To reiterate, even if a study is randomised, effect sizes are likely to be non-stationary as context differs (in space) changes (in time). Knowledge of these contextual drivers is essential to inform the conditions under which social technology can be transferred between settings.

Section 2 presented a list of stylised complex factors that may cause "complexity" in thinking about, evaluating and comparing PH interventions. System Dynamics appears to be powerful

enough to tackle such problems - if applied well. The purpose of this section is illustrative rather than applied; to provide a general overview of the rationale, approach and potential use of a systems approach in public health as an aid to discussion rather than prescription. Future work may seek to apply these approaches.

Causal mapping is an application of systems theory[8]. It is an intuitive approach that, at a minimum, may have the potential to structure our thinking about the complex nature of the social system that an intervention seeks to influence. The aims of causal mapping are three-fold: (i) to represent the problem; (ii) to serve as a basis for modelling, hypothesis formulation and simulation; and (iii) to allow for statistical inference - subject to data quality. Given length constraints a full description of these methods are not provided. We briefly outline each aim below, emphasising that the aim of statistical inference is a longer term objective. The main output of this approach in the short term may be to define more clearly areas of ignorance and essentially the value of information.

4.2.1 Representation: It is known in broad terms what the wider determinants of health are (e.g. social, legal, economic, environmental); and that there is likely to be a complex web of interactions between levels of a social systems, from individual to groups to communities and so forth. If we wish to understand how this affects the effect sizes of an intervention, we need to take the necessary next steps of defining these causal interactions more precisely, and representing uncertainty. "We" are always choosing between uncertain alternatives, and Decision Analysis allows us to do this more rationally.

To begin to uncover the relationships between drivers, the process of causal mapping often begins by collating expert opinion - much in the same way as "theories of change" or "logic model" might gather stakeholders to articulate programme level design – to articulate what the public health problem is, range of influences, key leverage points that an intervention targets and the causal pathways from intervention to impact. Causal mapping essentially attempts to quantify qualitative research with the purpose of explicating otherwise implicit causal relationships[9]. This aim is to develop a falsifiable structure to the problem: defining key variables; the nature and strength of interactions; and endogenise relevant contextual variables.

[8] Causal models are known by different names (e.g. graphical models). Essentially, it is the interaction between Graph theory and probability theory.

[9] While it is not uncommon for orthodox economics to think in this way (e.g. general versus partial equilibrium in macroeconomics; and rise of structural equation modelling in econometrics); systems approaches is really just a generalised exposition of the principles of modelling a complex system. Applications include systems engineering, information theory, operational research, artificial intelligence, and business and strategy.

The underlying premise in thinking about applying this to public health is a view that the causal drivers of public health are "systemic" in nature; and it's difficult for one intervention, in isolation, to make sustained difference to an aggregate outcome. Essentially, causal mapping can provide a nexus point for interdisciplinary research in public health.

4.2.2 Basis for modelling, hypothesis formulation, and simulation: There are a variety of techniques that ultimately share a common purpose but may have difference axiomatic and structural assumptions. Modelling choices include directions of causation (e.g. feedback versus acyclic); whether the interactions are "static" or "dynamic"; distributional assumptions of key variables and joint distributions of interactions; and impact of simultaneous interventions (and so forth). The point here is that there are a range of models to deal with a range of structural uncertainty: for instance, Bayesian networks (acyclical), Markov Networks (cylical) and Artificial Neural Networks (generalised and powerful).

Overall, the immediate application of causal mapping is not for inference (given lack of adequate data) but to transparently structuring of a complex problem, identifying areas of ignorance, and directing efforts to gather better information. In addition, there is a significant possibility for conducting simulation and scenario "what if" analysis, sensitising key variables and making projections onto common outcomes.

4.2.3 Application We are not aware of formal application within public health, though there is increasing interest in this approach (Trochim, 2006; Homer 2006). In behavioural change interventions there is an increasing focus on thinking about synergies between upstream interventions and downstream interventions. For instance, upstream interventions would be the importance of consistent public health messages, policies such legal changes (e.g. smoking ban); increasing the access and affordability of healthier lifestyles (e.g. green space and leisure facilities). Downstream interventions increasingly recognise the importance of culture feedback and support networks within local communities; hence behavioural interventions increasingly involving spouses, siblings, local groups etc. An approach of causal modelling would formalise this process and explicate implicit assumptions. It would be interesting to apply this to other areas of public health.

5. Key research area III: evaluation methods – study designs

Evaluators need to appraise four general dimensions: to what extent an intervention worked; cost; how it worked; and for whom. The latter two are typically overlooked in economic evaluation but are crucial to generalise lessons to other contexts. Again the running theme is that a combination of quantitative and qualitative methods may be required. This section emerges from an observation when randomisation is rejected (for ethical or feasibility concerns) study designs seem to default to using difference-in-difference approaches using imperfect controls, and highlighting ex post possible confounders. The purpose of the section is to briefly illustrate that there are a variety of methods to choose from; and in reference to evaluations at Glasgow (as introduced in Section 2 when illustrating the meaning of complexity) illustrate that are missed opportunities to adopt more rigorous estimates of effect sizes via quasi-experimental designs.

5.1 Estimating effects sizes: possible methods

As table 2 outlines there are three main categories of study designs, with a range of methods within each that can be considered. These are: randomised, quasi-experimental, and observational studies.

(i) Randomised control trials Randomisation is an approach, not a specific method, and provides the most credible estimate of effects. However, PH interventions are often population level or community based, consist of multiple components, and there may involve active contextual drivers. Classic RCT designs as defined in medicine (single standard intervention, with perfect controls and double blinding) are often considered unfeasible - although the ethical arguments are often overplayed. Nonetheless, alternative randomisation methods may still be possible. The recent MRC guidance (MRC, 2008) outlined the main variants of randomised approaches; cluster randomised control trials, randomisation to a waiting list, and stepped wedge designs; therefore, I do not repeat this.

ii) Quasi-experimental methods There are a number of other quasi-experimental designs that are less (not) acknowledged, but have considerable potential in public health. Done well these methods overcome ethical objections to orthodox randomisation, while approximately the results of RCTs. The most promising examples include “regression discontinuity” and “propensity score matching”.

Regression discontinuity[10] is where a programme/intervention is targeted at groups within the population based upon selection criteria defined by a composite variable. Given limited resources, a "cut-off point" is established to determine eligibility. This is common in welfare systems e.g. means tested benefits. On reflection it is apparent that a natural RCT may emerge around narrow margins of the cut-off point, as those on either side will not vary significantly, especially when random error is introduced. Two main conditions must be met. First, there needs to be a known relationship between the composite variable and the modifiable outcome measure in question. Second, the selection into programme must be known and made

Table 1: Categories of study design

Type	**Design**
Randomisation	Classic
	Clustered RCTs
	Step-Wise
	Waiting list
Quasi-experimental	Regression Discontinuity
	Propensity score matching
	Interrupted times series
	Pattern matching
Observational	Natural experiments
	Difference-indifference

on the basis of the composite variable. We are not aware of applications of Regression Discontinuity within UK (in any field). However, certain interventions lend themselves very well to it use.

For instance, a possible current application could be in programmes aimed at CVD prevention. Treatment guidelines recommend targeting patients most at risk of a cardiac event or stroke; estimated through a screening algorithm where patients with a probability of an event of 20% or greater (over the following 10 years) are given a mix of pharmaceutical and behavioural interventions. A natural RCT emerges if we sample heavily around the margins of the cut-off

[10] Regression Discontinuity has its origin in education and began in the 1970s, but dropped out of evaluation practise until recently.

point. We essentially measure the "discontinuity" (i.e. step change) in the post-intervention risk between the intervention and control group. In addition, there may be "interaction effects" given treatment success could vary contingent upon initial profile of the composite variable (e.g. starting risk score). If we have prior (pre-intervention) knowledge of the composite variable and outcome over the risk range (which could based upon previous longitudinal cohorts used to create the risk algorithm), then we also measure the interventions impacts across the entire range of risk above the cut-off also.

This approach could have been adopted in the economic evaluation of Keep Well, a CVD prevention programme piloted in Scotland since 2006. However, as the economic evaluation was included late in the programme/evaluation design, restricting our ability to influence the process. The economic evaluation is essentially simulating a range of scenarios of cost-effectiveness. This is done by asking "what if" questions regarding patient compliance in taking medications, assumptions around sustained behavioural change and then assessing the impact on reducing the risk of CVD and cost effectiveness. The literature on the medication is robust as expected (e.g. impact of statins); but the evidence base around behavioural interventions is extremely poor – especially regarding what type of intervention design (e.g. delivery mechanism(s), intensity, duration etc) works well for whom. The approach adopted is essentially a pre-trial model emphasising the value of information.

When any form of randomisation is not possible then the next best method may be propensity score matching (PSM). This is a response to the problem of imperfect control groups. Rather than simply using imperfect control adopting "difference-in-differences" measures of effect sizes and listing a range of exogenous influences that may be confounders, PSM seeks to adjust the control group itself. This is done by identifying relevant covariates that characterise the intervention and control population that are important in explaining the uptake and impact of an intervention. The control group is adjusted to "match" the intervention group as closely as possible across the relevant range of covariates – discarding poor matches if need be.

The key issue for PSM is which covariates to include? The crucial point is that there must be knowledge about relevant drivers that define the heterogeneity of the intervention and control groups. Therefore, in principle this provides a natural link to the previous section which advocated use of causal mapping to articulate more explicitly a theory of intervention. While not applied to

PSM (nor in reference to public health), there are recent authors who appear to be arguing in similar ways. For instance, Cartwright (2003) argues that by definition the selection of controls must be based upon a causal model. She goes onto to say that most often this is undertaken implicitly by economists, and as a result claims of counterfactuals can be spurious. If correct, the contention would be that the development of causal models and confidence in presumed counterfactuals go hand-in-hand.

Again an opportunity was missed in an evaluation recently completed at Glasgow. SHARP is a housing development and refurbishment programme that was instituted across different parts of Scotland. Randomisation was considered unethical, and drawing upon control groups an approach of difference-in-differences was adopted. However, the control groups were poorly defined and there was a huge amount of heterogeneity (different geographical areas, age groups, baseline living conditions etc). Attention to such factors was not rigorous and the sample size did not allow sufficient statistical power. In principle, a process of PSM could have been adopted which would have increased the sensitivity of the study design and confidence in the inference and generalisations made.

Less robust quasi-experimental methods but offering promise is natural experiments, abbreviate interrupted time series and pattern matching. They may have application especially if the intervention is at the population level and upstream (policy changes) where economic evaluations are very few. If the causal theory of intervention is well developed then this provides greater confidence regarding inference. Natural experiments are well known (Petticrew, 2003) therefore we briefly outline the others.

Abbreviate interrupted time series is where the population receives the intervention but its introduction and/or evaluation is staggered in time to corroborate effect sizes to attempt to account for population trends and reduce spurious inference due to confounding. Nonetheless, the quality of study designs can vary, and like any observational study lacking a randomised counterfactual, the risk is that exogenous variables may account for observed changes rather than the intervention. In the absence of classic controls, an attempt to account for confounders could be through the selection of a non equivalent dependent variable (NEDV). Essentially, this is a variable(s) that is independent of the study outcome measure, but which is also affected by the same set of contextual variables (and to the same degree). A simple example is the crackdown on drink driving using a

breathalyzer, with the presumption that drink driving as mainly during/shortly after pub opening hours. The dependent variable was causalities during pub opening hours and the NEDV was causalities while pubs were closed. The latter attempts to acts as a surrogate control for influences such as police crackdown on speeding, weather, safer cars and so forth. This design can be extended to multiple NEDV to increase the internal validity of inferences.

Very briefly, "pattern matching" responds to concerns of construct validity; concerned with whether empirical observations match the underlying theory. If an intervention demonstrates the theorised pattern of effects then this enhances the validity of inferences see Trochim (1985). This method requires a strong of theory, and again provides a case for attention to System Dynamics and Causal Modelling to increase rigour of hypothesis formulation and perhaps the use of simulation methods. These methods of abbreviated interrupted times series and pattern series can be used when control groups are not possible (though confidence in inference is reduced). Other disciplines such as natural science and engineering typically do not use control groups as the system itself is relatively well defined.

A natural question would be how good are these methods in practice? The validity of these methods has been tested in practice[11] through several "within study comparisons" (Cook et al 2008). This is where quasi-experimental results are compared with RCTs in the same intervention population. There are several robust within-study comparisons looking at the methods just outlined, primary within the fields of education, labour markets, and welfare. The results are impressive. The conclusions from Cook et al are that while randomisation is undoubtedly the most robust approach, quasi-experimental methods can work in social settings, though counterfactuals are likely to be contextually bound, and therefore generalisations of effect sizes are limited to similar environments. However, the precision of what context factors are important in generalisability are typically not formally defined. This is briefly considered in the next section. Given there are limited applications of these methods in public health research there is an opportunity to refine orthodox approaches and methods.

5.2 Choosing a method? Overall, the most appropriate method is that best suited to uncovering causality subject to constraints of ethics, feasibility and cost. In short, a plurality of

[11] Threats to validity include: delayed effects; maturation; reactivity; instrumentation and so forth.

methodological approaches may be the most appropriate when advocating general guidance: a matrix rather than a hierarchy (Petticrew et al, 2003).

6. Key research area IV: evaluation methods – generalisations:

Study designs essentially measure effect sizes. However, to inform transferability of a policy between settings requires an understanding of "how" an intervention worked and which groups benefited. In short, making generalisations is more difficult in PH interventions, where the design, implementation and context of a PH intervention can vary enormously. Heterogeneity in the population is less well defined than in HTA, and intervention effects can be driven by a range of contextual influences that can vary in space and over time. In practical terms, PH interventions that are similar in name can vary enormously in nature; and what work for one group may not work for other.

In recognition of this, the routine application of qualitative methods is likely to be required in PH research, and especially so if the intervention is small scale, the population is heterogeneous and there are consequent problems of statistical power[12]. In addition, qualitative methods could further shed light unpicking the effects of specific elements of programmatic interventions; and to verify (falsify) the presence (absence) of reinforcing (countervailing) influences.

An additional question is what to transfer. Traditionally there were two schools of thought in social interventions: the "fidelity" approach versus the "reinvention" approach. The former argues that the intervention must be replicated perfectly; where the latter argues against wholesale transfer and advocates for locally derived solutions. Increasingly, a third approach is proffered which says that as long as the causal mechanisms and steps in the intervention are maintained then local delivery can adapt to local capacity constraints.

Overall, the contention is that we need to define the "conditions for success" whereby a proven (through robust study designs) social technology can be generalised or transferred between different settings. These "conditions for success" would refer to the characteristics of the target population, institutional capacity (design and implementation), and key contextual drivers that interact with intervention effectiveness[13]. Essentially, these conditions should inform the

[12] "Reflexivity" typically raises eyebrows amongst more quantitative researchers, but it difficult to see how this is avoided. Qualitative research has various standards of evidence to reveal and minimise subjective bias.

[13] A dominant approach in qualitative evaluation called "Realistic Evaluation". The premise is that social reality and the mechanisms underlying an intervention effects are so context specific that generalising evaluation results is not

boundaries of a systematic review of a particular intervention/programme. That is, rather than perform a meta-analysis across PH in name where the reality there may be a diverse mix of actual designs, implementations, populations, and contexts; and more. This is the criticism often directed at Cochrane Reviews for instance.

Discussion

The main premise of the paper is that prioritisation is inevitable: decisions makers are choosing between competing public health interventions when allocating budget. However, that there is a lack of strategic focus: the maximand and criteria are unclear; the process of decision making is implicit, and ultimately concerns of opportunity cost and value for money are omitted.

The central focus of the paper is that Decision Analysis provides an approach that can capture the complex nature of public health interventions. Decisions are begin made anyway; so what form of analysis would allow for a better consideration of opportunity cost? Evaluation methods must be comprehensive enough to provide the information-set to help make allocatively efficient decisions. Evaluations need establish the cost effectiveness of an intervention, but to inform generalisations we also need to know how and for whom an intervention worked.

Areas for discussion:

Decision Analysis - Does Decision Analysis provide a valid approach to structure a complex problem (bearing in mind DA does not prescribe to a particular method or set of information inputs)?

In the short term, what are views as to applicability of Decision Analysis given the paucity of information to assess opportunity cost, and limitations of methods?

In the longer term, what level of sophistication could we attain; given the stylised features of complexity as outlined in section 2?

Common Outcomes - There is growing interest in the development of wider outcome measures, and these issues are well discussed elsewhere, hence its brief discussion here. However, given the

---

possible. The authors "abhor" concepts of counterfactuals and controls (Pawson). Such approaches and methods become exercises in complex description rather than evaluation. In particular, the strategic issue of Decision Analysis is side-stepped, and questions of opportunity costs and value for money ignored..

need to compare: interventions with impacts in short versus long term; different aspects of well-being; different populations and so forth, are we able to generate "a common measuring stick"? Eventually, alternative welfare (well-being) frameworks will come up against the same problems encountered in standard welfare economics – e.g. cardinal versus ordinal measurement; interpersonal comparisons; adaptive expectations etc.

Theory of Intervention - Do we need a more rigorous approach to the theory of intervention; and does the general approach of System Dynamics serve as a modelling basis (with its range of associated methods)?
Or, would this be a case of "Emperor's New Clothes"; hiding uncertainties through equally complex modelling methods? Thinking strategically, is the potential of Systems Dynamics and Causal Mapping limited to representing a problem; explicitly identify an array of complex influences; and serving as a basis for hypothesis formulation?

Study Designs - Is there sufficient awareness and use of robust study designs in public health interventions? For instance, is there still a requirement for the use of "difference-in-differences" within unadjusted imperfect controls when propensity score matching could be adopted?

Context - Do we (can we) have sufficient grasp of "context"; both as a confounder and effect modifier?

Generalisability - In reference to the above, what are the "contextual conditions for success" of transferring social technology; and what does this mean for conducting meta-analysis of PH interventions? Do we need context specific meta-analysis and how detailed?

PH evaluations can certainly be complicated, but it is hoped it is not all too complex to consider opportunity cost, allocative efficiency and therefore respond to the demand for uncovering value for money questions.

Thanks for your time!!

References

Bauman et al (1991) Reinventing fidelity: the transfer of social technology among setting American Journal of community Psychology Vol 19, no 4

Coast J et al (2008) Should the capability approach be applied in health economics? Health Econ 2008; 17(6):667-670

Cook, t et al (2008) Three Conditions under which Experiments and Observational Studies Produce Comparable Causal Estimates: New findings from within-study comparisons Journal of Policy and Management (in submission)

Cartwright 'Counterfactuals in economics: A commentary.' In Proceedings from INPC 2003:

Darizai (2008) "High quality care for all: NHS Next Stage Review final report"
Drummond et al (2007) Assessing the challenges of applying standard methods of economic evaluation to public health: CHE university of York

Ebrahim S et al (2006) Cochrane Multiple risk factor interventions for primary prevention of coronary heart disease (Review).

Homer et al (2006) PH System Dynamics Modeling for Public Health: Background and Opportunities American Journal of Public Health March 2006, Vol 96, No. 3

Hawe et al (2004), Complex interventions: how "out of control" can a randomised controlled trial be? BMJ 2004;328:1561-1563 (26 June),

Kelly et al (2006) Economic appraisal of public health interventions, Briefing paper health Development Agency NHS

Leyland et al (2003) Multilevel modelling and public health policy Journal of Public Health 31(4):pp. 267-274

Macpherson et al (2007) Theories of Change and Realistic Evaluation: Peas in a Pod or Apples and

Avril Blamey and Mhairi Mackenzie Evaluation 2007; 13; 439

McDaid et al (2007) the use of Economic Evaluation for public health interventions: desert of Oasis? A systematic review of the literature

MRC (2008) Developing and evaluating complex interventions: new guidance

Pawson et al (2004) – realistic evaluation, realist synthesis Realist synthesis: an introduction ESRC Research Methods Programme University of Manchester RMP Methods Paper

Petticrew et al (2003) horses for coursesEvidence, hierarchies, and typologies: horses for courses *Journal of Epidemiology and Community Health* 2003;57:527-529

Sheill (2007) What do we know about the value of public health & what do we need to know? NICE Health Population Health Intervention Research
http://www.nice2007.co.uk/AlanShiell.pdf
Rychetnik L, Frommer M, Hawe P, Shiell A. Criteria for evaluating evidence on public health interventions. *J Epidemiol Community Health* 2002;56: 119-27.

Sterman (2000). Business Dynamics: Systems thinking and modeling for a complex world, McGraw Hill.

Trochim (2006) PH Practical Challenges of Systems Thinking and Modeling in Public Health March 2006, Vol 96, No. 3 | American Journal of Public Health 538-546

Trochim (1985) Pattern Matching, Validity, and Conceptualization in Program Evaluation *Eval Rev* 1985; 9; 575

Wanless (2004) Securing good health for the whole population, HMT, 2004

PHI	Population	Design	Implementation	Context	Outcomes	Costs
Housing investment	***Simple:*** • Easily identified (deprived groups). • Paternal	***Simple:*** • Intervention(s), relatively discrete.	***Simple:*** •Single agency.	***Important:*** • Significant confounder.	***Complex:*** • Multiple diverse outcomes; • Long term impacts.	***Quite complex:*** •Mainly housing; •Some cross-sectoral implications.
CVD prevention	***Complex:*** • Hard to reach groups; • Requires voluntary engagement & motivation.	***Complex:*** • Multiple, interacting interventions.	***Complex:*** •Multiple & diverse stakeholders; •Patient engagement & long term compliance; •Committed staff, • Multiple skill sets, • Institutional coordination.	***Important:*** • Significant confounder, • Determinant of design & implementation; • Interactions with effect.	***Simple:*** • CVD risk reduction; • Short term impacts.	***Quite complex:*** •Mainly NHS; •Other welfare sectors; •Various institutional levels including central govt, health boards; NGOs.
Regeneration	***Simple:*** •Easily identified (deprived groups); •Mainly paternal.	***Complex:*** • Within transformation areas; • Multiple, interacting interventions.	***Complex:*** • Different interventions (and intensity) for different areas; • Multiple stakeholders; • Neighbourhood participation; • Long term commitment; • Institutional coordination.	***Important:*** • Significant confounder; • Determinant of design & implementation; • Interactions with effect.	***Complex:*** • Multiple diverse outcomes; • Long term impacts.	***Quite complex:*** •Mainly housing; •Regeneration (which sector?).