Working Paper # 2

'Good' evidence for improved policy making: from hierarchies to appropriateness

Sudeepa Abeysinghe, Justin Parkhurst

June 2013

London School of Hygiene and Tropical Medicine

GRIP-Health Programme

www.lshtm.ac.uk/groups/griphealth

GETTING RESEARCH INTO HEALTH POLICY AND PRACTICE

GRIPHEALTH

Summary

Within the field of public health, and increasingly across other areas of social policy, there are widespread calls to increase or improve the use of evidence for policy making. Often these calls rest on an assumption that improved evidence utilisation will be a more efficient or effective means of achieving social goals. Yet, a clear elucidation of what can be considered 'good evidence' for policy use is rarely articulated. Many of the current discussions of best practice in the health policy sector derive from the evidence-based medicine (EBM) movement, embracing the 'hierarchy of evidence' in framing the selection of evidence – a hierarchy that places experimental trials as preeminent in terms of methodological quality. However, there are a number of difficulties associated with applying EBM methods of grading evidence onto policy making. Numerous public health authors have noted that the hierarchy of evidence is a judgement of quality specifically developed for measuring intervention *effectiveness*, and as such it cannot address other important health policy considerations such as affordability, salience, or public acceptability (Petticrew and Roberts, 2003).

Social scientists and philosophers of knowledge have illustrated other problems in the direct application of the hierarchy of evidence to guide policy. Complex or structural interventions are often not conducive to experimental methods, and as such, a focus on evidence derived from randomised trials may shift policy attention away from broader structural issues (such as addressing the social determinants of health (Solar and Irwin, 2007)), to disease treatment or single element interventions. Social and behavioural interventions also present external validity problems to experimental methods and meta-analyses, as the mechanisms by which an intervention works in one social context may be very different or produce different results elsewhere (Cartwright, 2011). In these cases, policy makers may be better advised to look for evidence about the mechanism of effect, and evidence of local contextual features (Pawson et al., 2005).

We argue that rather than adhering to a single hierarchy of evidence to judge what constitutes 'good' evidence for policy, it is more useful to examine evidence through the lens of *appropriateness*. It is important to utilise evidence to improve policy outcomes, yet the form of that evidence should vary depending on the multiple decision criteria at stake. Policy makers must therefore start by articulating their decision criteria in relation to a given problem or policy, so that the appropriate forms of evidence can be drawn on – from both epidemiological and clinical experiments (e.g. for questions of treatment effect), as well as from social scientific, social epidemiological, and multidisciplinary sources (e.g. for questions of complex causality, acceptability, human rights, etc.). Following this selection of types of evidence on the basis of appropriateness, the rigour and quality of the research can be assessed according to the evidentiary best practice standards of the discipline within which the evidence was produced. This approach speaks to calls to improve the use of evidence through ensuring rigour and methodological quality, yet recognises that good evidence is dictated by specific public health or social policy goals.

1 'Good' evidence for improved policy making

Introduction

The introduction of the concept of evidence-based policy has marked an important shift in policy processes. The health sector has particularly embraced this idea, in part because of the easy analogy with the evidence-based medicine movement (EBM), which has driven many of the current ways of using evidence within policy (Cookson, 2005, Berridge and Stanton, 1999). It is now generally acknowledged that the using research evidence to inform policy making can produce more efficacious results.

However, the use of evidence within policy is as yet an unclear process. Previously, it had been thought that policy makers could draw directly upon research evidence where necessary, or conversely, that researchers could present and gear research in a way that optimises its adoption by policy makers. Early work on 'knowledge transfer' implied, as the term suggests, that the process was one of simply transferring the knowledge produced by researchers in a policy-useful format. Following from this, a wide range of efforts have been undertaken to increase the linkages between researchers (or their research findings) and decision makers (see Lavis et al., 2003, Mitton et al., 2007 for summeries of the knowledge translation literature). Further explorations suggested that 'bridging the gap' between the worlds of research and policy is not always straight-forward (see also Greenhalgh and Wieringa, 2011 for a critique of the concept). It has been noted, moreover, that the linear understanding of the evidence-to-policy process does not adequately account for the complexities and political nature of policy making (Bowen and Zwi, 2005). There are many factors, inherent within the political process of policy making, which might complicate the use of research evidence. Similarly, it is also important to note that research is often not produced in a way that is readily consumable for policy actors (Lavis, 2006, Lavis et al., 2003). The goals of academic researchers do not necessarily translate directly to the goals of policy makers. An additional problem lies in understanding which pieces of evidence (i.e. bodies of literature or particular studies) might be useful for any particular policy problem. Such issues frame the focus of the following discussion.

This working paper summarises existing ideas surrounding the good use of evidence. It focuses upon the current primacy of models which emphasise techniques drawn from evidence-based medicine (EBM), and the 'hierarchy of evidence' that EBM relies upon. It is shown that existing models of best practice tend to emphasise certain methodological elements (which favour experimental approaches) as critical to the ranking of quality evidence. The paper then explores critical voices from within public health, but also from sociologists and philosophers of science, on the issue of evidence use. These commentators point out that the forms of evidence highlighted as superior by the hierarchy of evidence are based upon a narrow view of methodological quality, specifically designed to address questions of intervention effect, and do not help to answer many questions which have social, cultural, or political dimensions. Instead, other bodies of evidence may be more appropriate to answering those questions, each with their own criteria for quality.

This paper attempts to explain some ways in which the use of evidence can be improved, taking into account existing critiques, but in a way that is practical and useful for public health planners. It proposes that the best use of evidence in decision-making does not simply focus upon quality as judged by the hierarchy of evidence. Rather, it is more useful to judge the appropriateness of the evidence type in respect to the considerations of the decision-maker. We suggest that the first step in the appropriate utilisation of evidence should therefore be the explicit articulation of policy objectives and decision-making criteria – both the biomedical and the broader social-political or economic concerns linked to a health policy decision. Following this, evidence should be selected on the basis of its appropriateness to the particular policy objectives, allowing for a more accurate matching of evidence to policy needs. Only after this should the evidence be assessed in terms of methodological rigour, based upon the type of evidence selected.

'Best practice' as a hierarchy of evidence

Current approaches to the use of evidence in policy have been drawn from the tradition of evidence-based medicine (EBM). The EBM movement highlights the importance of using evidence (particularly epidemiological evidence) to shape clinical decision-making (Canadian Taskforce on the Periodic Health Examination, 1994, Evidence-Based Medicine Working Group, 1992), and has been geared towards, and best applied to, questions of treatment efficacy.

The dominant model for assessing evidence within EBM is drawn from the 'hierarchy of evidence' from the natural sciences. This hierarchy sets out the process though which research evidence can be evaluated. Forms of evidence that most adhere to the ideals of experimental conditions (as given in the natural sciences) are set at the 'top' of the hierarchy. These are methods which display key characteristics which include: large and representative sample size, control for experimenter and participant bias (often in the form of blinding, or preferably, double-blinding); control for external variables (i.e. studying the problem within a laboratory environment and/or use of a control arm to exclude confounding variables); the study of a singular experimental variable (to determine direct cause- effect relationships); and value-neutrality (i.e. the idea that the researcher must not be intent on a certain outcome, or let their subjective ideas impact on the research process) (Merton, 1973).

It is understood that for clinical interventions, these factors are best constituted in the form of the Randomised Controlled Trial (RCT). Randomisation is understood to overcome the problem of confounding, by ensuring that any significant difference observed between subject groups is only due to the experimental variable/intervention. Experimental trials also attempt to minimise bias from either the researcher (particularly in double-blind conditions, where the research themselves do not know which research subjects were treated and which were the control group) and subject bias (again through blinding, since

research subject may tend to unconsciously behave in certain ways to please researchers or otherwise skew research results) (Chalmers et al., 1981). The use of a placebo (i.e. in the control group of subjects) also lets researchers account for the placebo effect, or the extent to which simply being studied achieves a change in behaviour or state.

Non-experimental methods - such as case studies, observational data, or case-controlled studies – are seen as less useful forms of intervention research, due to their inability to control for confounding variables, and the greater potential for bias to be introduced as some stage in the research protocol. However, these forms of research can also be more or less rigorous, depending for example upon sample size, representativeness, and other qualities of the methods employed (Borgerson, 2009).

The norms of good scientific method, as illustrated above, define which types of research are considered 'best' in relation to the hierarchy of evidence, and what 'rigour' means in the context of these types of research. The way in which the hierarchy is described can differ slightly between organisations and commentators. However a simplified hierarchy consists of the following:

1. Systematic reviews and meta-analysis of Randomised Controlled Trials (RCTs)
2. RCTs with definitive results (large and well-conducted studies)
3. RCTs with non-definitive results (including smaller RCTs)
4. Cohort studies
5. Case control studies
6. Case studies
7. Expert opinion

(see Nutley et al., 2012 for variations on the hierarchy)

Though these categories are somewhat variable, all representations emphasise large, randomised and well-controlled trials as the gold-standard of research. For example, in the UK, the National Institute of Health and Clinical Excellence (NICE) provides guidelines to the National Health Service, and has its own hierarchy of evidence to grade the quality of evidence for its recommendations (see for example NICE, 2004). This is integrated with cost-effectiveness studies to produce recommendations, which are awarded 'grades' depending upon the strength of their sources, with 'A' recommendations being based directly on RTCs or meta-analyses of RTCs to 'D' recommendations based upon expert opinion or inferences from upper-level studies (NICE, 2005: 11.5).

There are also a number of other bodies using similar hierarchies to guide health policy and practice. The GRADE (Grading of Recommendations, Assessment, Development and Evaluation) working group, for example, is an international body that aims to develop a more universal mechanism to grade evidence of health interventions to develop recommendations. GRADE (2013) evaluates biomedical evidence upon the basis of risk,

burden, and cost of intervention. This brings in some non-biomedical factors (e.g. cost of intervention), but again the initial approach is still to judge evidence from RCTs as high quality, from observational data as low quality, and other methods as very low quality.

Similarly, the Strength of Recommendations Taxonomy (SORT) (Ebell et al., 2004), formed by a consortium of family medicine practitioners and academics, is aimed at helping physicians navigate the process of EBM by assessing the quality, quantity and constitution of evidence based upon EBM hierarchies of evidence. A further example of the way in which EBM techniques has been formalised is through the Centre for Evidence Based Medicine (CEBM), run through Oxford University, which is designed to aid physicians, researchers and patients to understand EBM approaches (2013). CEBM guidelines go a long way in qualifying the use of these approaches (i.e. in explicitly cautioning that whole-population based approaches do not straight-forwardly indicate what might be best for an individual patient). Despite this, common to all of these approaches is the way in which the methodological superiority of experimental evidence, and hierarchies of evidence formed around this, are taken for granted (see also Annex 1 of Nutley et al., 2012 for other examples of evidentiary management bodies).

What is clear in these formulations of 'best' evidence is the fact that certain research methodologies are placed above others. Particularly privileged are randomised trials, and combinations of multiple randomised trials which show consistent effects. This way of evaluating evidence is useful in so much as it allows policy makers to easily sift through large amounts of research and identify the most rigorous pieces (Cook et al., 1997, Mulrow, 1994). However, many commentators have also pointed out potential flaws in this technique, for example where small studies, conducted in particular contexts, are combined in a way that skews results (Black, 2001).

More fundamentally, evidence evaluation techniques which are based upon EBM take for granted that evidence can be assessed in relation to its methodological 'quality', as defined by the norms of the natural sciences. These methods presuppose that causal mechanisms are constant over place and time. This also assumes that taking a research problem out of its context is the best way of understanding it (and this assumption, as we will show below, that can often be problematic).

The establishment of these rankings of evidence have typically grown out of a concern for ensuring that practice – particularly clinical practice – follows the best available evidence of effectiveness. However, these hierarchies are increasingly being applied in policy circles. The shift in terminology from evidence based medicine to evidence based policy has equally seen attempts to call for policy decisions to also apply such hierarchies of evidence to their decision making. Yet this raises a number of critical questions.

The importance of non-clinical outcomes in policy decisions

Policy making often involves deciding between competing sets of decision criteria. Health policies may be decided on the evidence of clinical effect of an intervention, but decision makers equally may wish to consider the social acceptability of that intervention, or the impact it will have not just on morbidity and mortality outcomes, but on other socially valued concerns, such as equity, justice or human rights. Many health policy decisions are not simply about clinical and biomedical interventions, but may involve social and organisational interventions for which these hierarchies were not originally developed.

Even within the field of public health there have been voices pointing to the misuse of evidence hierarchies to inappropriate questions (Booth, 2010, Petticrew and Roberts, 2003). As Glaszou and colleagues explain, "different types of question require different types of evidence" (2004: 39). For most policy making situations, the different types of questions go beyond clinical and immediate health related issues, to involve areas of social, political or economic concern. RCTs are not always useful for questions that do not speak directly to clinical efficacy, and have been criticised for being applied uncritically, even within the biomedical sciences (for example, in investigating disease aetiology rather than treatment options) (Glasziou et al., 2004, Green and Glasgow, 2006). The external validity of many RCTs, indicating the usefulness of the research in the context of different patient demographics, is often not well-articulated (Rothwell, 2005). Further, causality is often a complex process, and RCTs are not necessarily helpful in situations where multiple causal factors might be implicated (Victoria et al., 2004). As such, calls for methodological aptness (Pettigrew, 2003), and a context-based selection of evidence (Boaz and Ashby, 2003, Dobrow et al., 2004) are now coming to the forefront.

Political scientists have long noted the multiple competing values and issues around which policy decisions are made, pointing to the need for policy makers to consider multiple bodies of evidence, including evidence surrounding social values and norms. These will not come from experimental methods. Rather, such evidence will come from methods which seek to understand (rather than seek to control for) the social context (Petticrew and Roberts, 2003, Bowen and Zwi, 2005). Policy interventions with social components, or which seek out social change, need to look at forms of research which provide information on the social (rather than natural) world.

Hierarchies of intervention effectiveness do not well-inform many important policy goals

The sociology of health highlights the fact that ill-health is often structured by gradients of socio-economic status (Wilkinson, 2002, Wilkinson and Marmot, 2003), gender (Courtenay, 2000, Doyal, 2000), geographical location (Haynes and Gale, 2000), or other social variables. If public health officials ultimately strive to alleviate ill-health, or identify the cause of ill-health, it may be useful for them to utilise evidence from research on social variables, much

of which is not experimental in nature. For example, diabetes mellitus is enduring as an important chronic health problem in countries which have undergone the epidemiological transition into chronic disease prevalence. In accounting for diabetes through policy, decision-makers might find it useful to seek out clinical evidence on risk factors and treatments. However, policy making might also seek to target at-risk populations and communities. In order to do this, research that explores the social distribution of diabetes can help policy makers understand the problem further and provide as much, if not more, useful evidence to inform decisions as experiments of interventions to treat or prevent diabetes. These may rigorously test the effectiveness of specific interventions, but do not speak to the socio-political considerations of relevance. Clinical effectiveness evidence may also unduly focus policy makers on treatment over prevention, particularly when causes are complex and socially rooted. Looking at the literature surrounding the social gradient of diabetes illustrates that its incidence is structured by sex, ethnicity, socio-economic status and other social factors (McKinlay and Marceau, 2000, Young et al., 1990). This type of research evidence might therefore be more important to guide the management of this disease in the long-term.

As suggested within the sociology of scientific knowledge, the existing hierarchies of evidence are based upon an understanding of health and illness as purely biological phenomena (Goldenberg, 2006). As a result, they highlight studies that seek out biological universals (that is, in seeing all bodies as fundamentally the same, they try and omit confounding variables in the study of biological processes). However, even if though biochemistry and anatomy may be fairly consistent, human behaviour, socio-cultural values, and social and political structures are widely variable. As the sociology of health and illness illustrates, there are many social factors that impact upon health and healthcare. For example, healthcare generally occurs within the confines of professional and institutional structures. Understanding these structures can therefore be useful to understanding the way in which health outputs can be optimised.

A simple example of health service management helps to illustrate: If a Ministry of Health wants to improve the flow of patients within public hospital emergency rooms, they can think of several ways in which this can be achieved (for example, increasing the number of emergency room beds, modifying the way in which patients are triaged etc.), all with distinct economic and political advantages and disadvantages. When looking at this question, experimental forms of research are feasible – one could randomly allocate some hospitals to have more beds, others to have different triage processes, and others as controls. Yet the complexity of the causal mechanism may mean that such experiments on single variables may not adequately address the policy problem. Experiments varying single components of a complex system may be less useful than efforts aimed to better understand the structure and organisation of emergency rooms as a systemic whole. One way to do this could be through observational research. For example observational studies look at the way in which patients 'flow' through hospital systems. For instance, Nugus and

colleagues find that efficient flow of patients depends on many factors, including the mobilisation of personal and professional influence, hospital management structures, as well as ways in which staff on non-emergency wards perceive and/or guard the 'space' left on their ward (Nugus and Braithwaite, 2010, Nugus et al., 2009, Nugus et al., 2010). Similarly, mathematical models of patient flow, or interview data on health workers' experience of patient flow in the A&E may help make the situation more clear to policy makers. The intervention decision may therefore be based on a tailored approach based on understanding of system dynamics within a given hospital setting, rather than application of a tested and 'proven effective' universal approach. While the types of research forwarded by hierarchies of evidence are potentially helpful, forms of evidence that seek to understand (rather than control for) social context may be equally (or more) useful.

Social norms and behaviours are integral to illness and to the management of illness (Helman and Helman, 2007). Many health policies must therefore take into account aspects of social or behavioural change to achieve optimal results. This provides another challenge to reductionist applications of a hierarchy of evidence, which value experimental trials (which typically are of single interventions) with an expected generalisable causal effect. In social interventions, often the mechanism of effect is contextually determined, and, as such, the mechanism through which an intervention works in one place, or population, or time, may be very different elsewhere (Cartwright, 2011, Pawson and Tilley, 1997). For example, increasingly the HIV prevention field has been focussing upon structural interventions to reduce behavioural HIV risk (Auerbach et al., 2011, Gupta et al., 2008), with recent discussions on whether financially based interventions – such as cash transfers or access to credit (e.g. microcredit loans) - are 'effective' for preventing HIV (Baird et al., 2012, Kohler and Thornton, 2010, Medlin and De Walque, 2008, Hall, 2006, Pronyk et al., 2006). However, the social nature of sexual risk behaviour (and any links is has to access to financial resources) means that a financial intervention showing an impact in one area may work in very different ways elsewhere. So while an intervention that provides financial assistance may lead to reduced HIV-related risk when given to poorer women who rely on transactional sex to make ends meet, the exact same intervention may increase risk taking in another setting - for instance if given to women who never relied on transactional sex, but who end up using the funds to travel and as a result end up engaging in wider sexual networking. Similarly provision of HIV/AIDS information has been studied as if there is a single mechanism through which information may affect behaviour, yet an information campaign that inspires fear in one setting to achieve behaviour change might inspire laughter or disgust in another, working (or not working) through very different mechanisms of effect.

Meta-analysis is often held up to be at the top of the evidence hierarchy, yet the above example illustrates how unfit it can be for the purpose of guiding policy action if the mechanism of effect of an intervention changes according to local contextual factors. If a meta-analysis combined trials of cash transfer interventions for HIV prevention and included

a population for whom it averted transactional sex alongside a population for whom it promoted wider sexual networking, the final conclusion might erroneously be 'cash transfers show flat (or conflicting) results'. Yet a more accurate (and more useful) conclusion might be that 'cash transfers work for some groups in some contexts, and do not work for other groups in other contexts'. To draw this conclusion, however, requires different evidence – not just an increasingly large sample on whom the intervention has been trialled, but 'realistic evaluation' evidence (or a 'realist' review) that investigates how social context affects the mechanism of intervention to achieve an outcome or impact (Pawson et al., 2005, Pawson and Tilley, 1997). This might include ethnographic evidence or in-depth interviewing in target communities in this example, for instance, in addition to any trial of effectiveness (cf. Bonnell et al. 2012 for an attempt to integrate these approaches). As Nancy Cartwright has explained "[f]or policy and practice we do not need to know 'it works somewhere'. We need evidence for 'it-will-work-for-us'." (Cartwright, 2011: 1401). Context specific, and therefore inherently social, factors can therefore be seen as worthy and necessary of study – and a body of evidence particularly necessary to inform policies of this nature.

From a hierarchy to *appropriateness*

For these reasons detailed above, public health (and other social policy) decision makers may find that a simple application of the hierarchy of evidence does not best serve their policy goals. In order to best apply evidence to policy, decision makers need to understand both the multiple decision criteria on which the policy decision is based, as well as the nature of the interventions they aim to implement to achieve their policy goals. If a proposed intervention has purely clinical aspects, and the only policy criteria at stake is morbidity, mortality, or cost-effective criteria, then the evidentiary best practice might indeed be to follow hierarchies of evidence from epidemiology and clinical medicine. If aspects of the health problem or proposed solution are social or behavioural, or if other social outcomes are an important part of the policy decision, then different sets of evidence can be sought out.

In order for the appropriate evidence to be chosen, therefore, policy makers also need to play an active role. The underlying goals and premises of the policy need to be well-established before the evidence can be chosen. This includes an explicit articulation of which factors are considered to take primacy in making decision (e.g. to what extent do economic considerations overweigh the benefits of the proposed intervention, to what extent does positive impact in a small sub-population justify a large-scale or costly change etc). Pinpointing the goals of the policy as explicitly and narrowly as possible ensures that instead of being daunted by large amounts of varied research, a narrower, appropriate and specific set of research can be accessed.

What is required, then, is both an explicit understanding of the nature of the policy question (what is it that is needed from the evidence), and a more nuanced understanding of what might constitute 'good' evidence for particular policy concerns.

Appropriate, but rigorous, evidence

Once the appropriate evidence has been determined, the various forms of evidence can be assessed for quality. The validity and rigour of different forms of evidence is established by different methodological criteria. As noted, current hierarchies of evidence look at the idea of good research through a narrow perspective which emphasise qualities that are appropriate for the research of context-free and universal biological or physical properties with an expected direct causal effect. This way of understanding rigour is useful for most clinical evidence of biological or surgical actions and treatments. It is also applicable for some epidemiological research, which seeks to understand risk factors, or the success of population-level interventions. In contrast, as demonstrated above, policy makers may also need to access those forms of evidence which derive from the very different realities of the social and political world.

Each research or methodological tradition (i.e. experiments, interviews, observations etc.) is underpinned by its own standards of quality and validity. Awareness of each is important because different research methodologies seek to understand different parts of the process of health and healthcare. There is a fundamental difference in trying to understand the biological, the individual, the social, the economic and the political, and each produce research through very different lenses. Since this is the case, methodological traditions are accompanied by different research protocols and different forms of rigour.

For example, survey research may be useful when evaluating the opinions of communities around social acceptability. When looking at survey research, the quality of evidence should be evaluated in a way specific to that research tradition. This would include an assessment of statistical representativeness, including the sample size and variation. Rigorous research in the context of surveys would also include studies which exhibit internal validity – the questions asked in the survey actually measure and reflect the aims of the research – and also external validity – that the results are generalisable to the target research population, achieved by making sure that the survey instrument is properly representative. There are other ways in which a survey can maximise reliability, for example through triangulation, in asking research subjects several questions that are aimed to provide data on a singular research question (i.e. to make sure that, even when the same subject/question is broached through different wording or emphasis, the results remain consistent). Controlling for the conditions in which the survey is performed will also help to maximise validity. For example, research subjects should ideally by surveyed in similar environments (i.e. all at their homes, all at their GPs surgery etc.), in identical ways (in terms of the process of administering the

survey, the emphasis or tone of voice in the case of verbal surveys etc.). These mechanisms are set out to control the influence of external factors wherever possible.

Observational or ethnographic research may be useful to policy makers in understanding the cultural context that surrounds a certain policy room (such as the A&E example above), or to access the perspectives of a small but important community of people. So, for instance, if a policy problem calls for a better understanding of the way in which breast cancer patients make sense of their diagnosis and prognosis (for example, in order to produce policy that bettered the experience of such patients), one way in which this could be done is through observing the communication of diagnosis (see Gross, 2009 in the context of brain cancer diagnosis). Observational and ethnographic studies emphasise the importance of understanding processes through the perspective of key participants (Hammersley and Atkinson, 1989). Since context and meaning are so strongly tied to this research tradition, these are emphasised in the understanding of evidentiary rigour for these methods. High quality observational and ethnographic research is signified by the researchers' immersion in the research context and the ability of the research to gain insider insight into processes. One way in which the researcher can know this is achieved is through feeding back their findings to the research participants, as a method of seeing if the account of the research is valid to those involved. Another criteria of rigour and validity in terms of this research tradition is the idea of reflexivity – that is, since the researcher is immersed in the context, it is important for the researcher to be able to explicate their own values or viewpoints may have impacted upon their understanding of the process (Davies, 2008). Unlike a survey technique, then (which attempts to minimise external influences in some sense), it is acknowledged that the researcher is necessarily 'close' to the process, and validity is accounted in relation to the ability of the researcher to accurately articulate the process.

Interviewing methodologies, on the other hand, occupy a broad spectrum between survey and ethnographic approaches. Interviewing techniques can be useful for policy makers where the in-depth opinions and viewpoints of a small number of individuals is useful. For instance, when looking at the breast cancer diagnosis as given above, it might be appropriate to interview the oncologists or GPs involved in the communication of diagnosis to try and access their perspective on the process and impact on patients. For the example of cash-transfers for HIV prevention above, interviews might be needed to identify how access to cash affected specific risk behaviours, and for which sub-groups the intervention appeared to be more or less successful. These insights could be produced through one of many forms of interviewing. These range from structured techniques (where questions are set, and the same questions are asked of each interviewee) to semi-structured interviews (where interviewers are bound to the set of core problems, but may ask slightly different questions to different participants in order to access data surrounding this core set of problems) to unstructured interviews (where the course of the interview is tailored to each participant, not pre-determined, and allowed to follow the course of the conversation
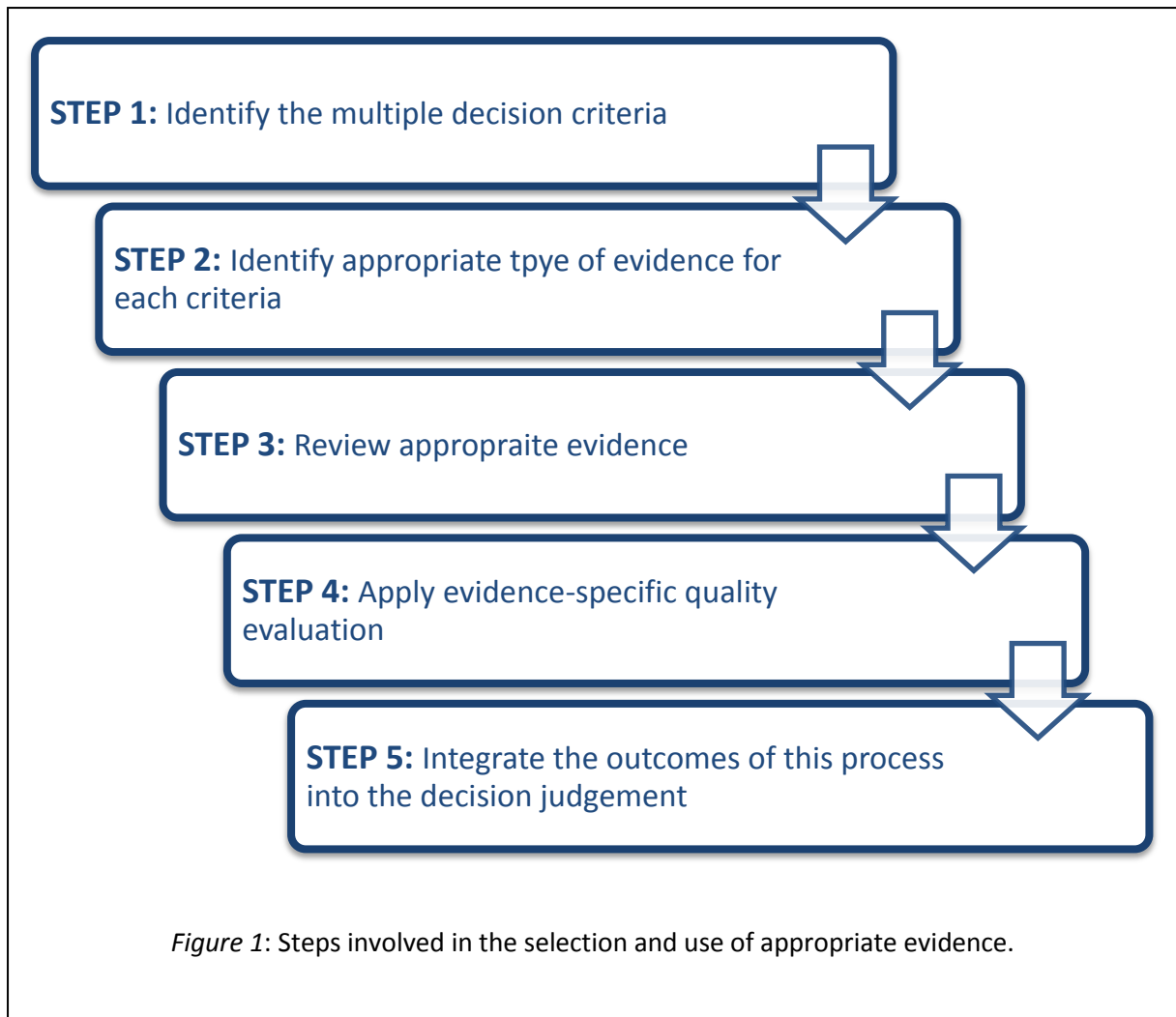
between the interviewer and interviewee) (Silverman, 2004, 2009). Since these forms of interviewing are diverse, methodological rigour is assessed slightly differently in each case. In the case of strongly structured interviews, ideas of validity are closer to that of survey methods (i.e. in the context of reliability and maintaining a regimented process). Strongly unstructured interviews emphasise forms of validity more close to ethnography (i.e. how well does the data reflect the experiences of the research participants). Other forms of rigour in the context of interviews include the idea of 'saturation' – rigorous interview protocols conduct interviews until no 'new' data appears (Bowen, 2008). Equally, conclusions must be based upon multiple interviews, and not simply extrapolated from a few cases.

Ultimately, when selecting evidence, what is essential is for decision makers to firstly identify the types of information they need on which to base their decision (their decision criteria) after which, the appropriate evidence can be judged and evaluated. Each research tradition comes with its own criteria for establishing 'rigour'. Once the appropriate evidence base is selected, these assessments of rigour can be applied as according to the criteria set out by research of that tradition.

Conclusion

This paper has been developed within a research programme concerned with improving the use of research evidence in health policy. However, to understand how to do this, a key question revolves around what 'good' evidence for decision making looks like. The multiple social aspects of any health problem or intervention are integral to the management of illness and achieving public health goals. Due to the fundamentally social nature of health and illness, and the contextual realities of healthcare and health policy, notions of evidentiary validity derived from clinical medicine and the evidence-based medicine movement do not necessarily extend to all the questions posed in the formation of effective public health policy. The Western medical ideal sees research and causality as divorced from social consequences. However, on the contrary, health and illness are fundamentally socially embedded – and questions surrounding both the origins of health problems, as well as the management of health conditions are typically socio-political in nature.

We argue that 'good' evidence should not simply be equated to a particular position within the hierarchy of evidence of the natural sciences, which specifically relates to effectiveness studies. Rather, we argue for a conceptualisation which sees good evidence for policy as that evidence which is *appropriate* to the multiple decision criteria being considered. Once these decision criteria are elucidated, and evidence bodies identified, then the quality and rigour of each evidence type can further be evaluated before the ultimate policy judgement is made. The figure below attempts to provide a simple schematic for this process:

STEP 1: Identify the multiple decision criteria

STEP 2: Identify appropriate tpye of evidence for each criteria

STEP 3: Review appropraite evidence

STEP 4: Apply evidence-specific quality evaluation

STEP 5: Integrate the outcomes of this process into the decision judgement

Figure 1: Steps involved in the selection and use of appropriate evidence.

There are almost no decisions at a political level that simply require an analysis of epidemiological or clinical evidence. Every decision has opportunity costs, and most health issues touch on a range of important concerns beyond morbidity and mortality – such as economic impact (not just cost-effectiveness), fairness and equality, solidarity and justice, or human rights. Many health interventions further involve social norms and behaviours, or government actions over which the population may have moral or ideological concerns (such as views about state control versus individual freedom, or the 'right' and 'wrong' way to behave). The vast majority of these issues cannot, and should not, be addressed with evidence that easily fits into a single hierarchy. For public health actors to achieve their policy goals – goals such as improvements in population health, reductions in avoidable morbidity and mortality, and decreases in health inequalities – they must ensure not only that they use evidence to guide their decisions, but that they use the right evidence to do so.

REFERENCES

AUERBACH, J. D., PARKHURST, J. O. & CÁCERES, C. 2011. Addressing social drivers of HIV/AIDS for the long-term response: conceptual and methodological considerations *Global Public Health,* 6**,** S293-S209.

BAIRD, S. J., GARFEIN, R. S., MCINTOSH, C. T. & ÖZLER, B. 2012. Effect of a cash transfer programme for schooling on prevalence of HIV and herpes simplex type 2 in Malawi: a cluster randomised trial. *The Lancet*. 379, 1320-1329

BERRIDGE, V. & STANTON, J. 1999. Science and policy: historical insights. *Social Science & Medicine,* 49**,** 1133-1138.

BLACK, N. 2001. Evidence based policy: proceed with care. *BMJ: British Medical Journal,* 323**,** 275.

BOAZ, A. & ASHBY, D. 2003. *Fit for Purpose?: Assessing Research Quality for Evidence Based Policy and Practice,* London, ESRC UK Centre for Evidence Based Policy and Practice

BONELL, C., FLETCHER, A., MORTON, M., LORENC, T. & MOORE, L. 2012. Realist randomised controlled trials: A new approach to evaluating complex public health interventions. *Social Science & Medicine,* 75**,** 2299-2306.

BOOTH, A. 2010. On hierarchies, malarkeys and anarchies of evidence. *Health Information & Libraries Journal,* 27**,** 84-88.

BORGERSON, K. 2009. Valuing evidence: bias and the evidence hierarchy of evidence-based medicine. *Perspectives in Biology and Medicine,* 52**,** 218-233.

BOWEN, G. A. 2008. Naturalistic inquiry and the saturation concept: a research note. *Qualitative Research,* 8**,** 137-152.

BOWEN, S. & ZWI, A. B. 2005. Pathways to 'Evidence-Informed' Policy and Practice: A Framework for Action. *PLoS Medicine,* 2**,** 600-605.

CANADIAN TASKFORCE ON THE PERIODIC HEALTH EXAMINATION 1994. *The Canadian Guide to Clinical Preventative Medicine.* Ottowa: Canada Communication Group.

CARTWRIGHT, N. 2011. A philosopher's view of the long road from RCTs to effectiveness. *The Lancet,* 377**,** 1400-1401.

CEBM. 2013. *Centre for Evidence Based Medicine* [Online]. University of Oxford. Available: http://www.cebm.net/ [Accessed 01/05/13.]

CHALMERS, T. C., SMITH, H., BLACKBURN, B., SILVERMAN, B., SCHROEDER, B., REITMAN, D. & AMBROZ, A. 1981. A method for assessing the quality of a randomized control trial. *Controlled Clinical Trials,* 2**,** 31-49.

COOK, D. J., MULROW, C. D. & HAYNES, R. B. 1997. Systematic reviews: synthesis of best evidence for clinical decisions. *Annals of Internal Medicine,* 126**,** 376-380.

COOKSON, R. 2005. Evidence-based policy making in health care: what it is and what it isn't. *J Health Serv Res Policy,* 10**,** 118-121.

COURTENAY, W. H. 2000. Constructions of masculinity and their influence on men's well-being: a theory of gender and health. *Social Science & Medicine,* 50**,** 1385-1402.

DAVIES, C. A. 2008. *Reflexive Ethnography: A guide to Researching Selves and Others*, Routledge.

DAVIES, P. 2000. The relevance of systematic reviews to educational policy and practice. *Oxford Review of Education,* 26**,** 365-378.

DOBROW, M. J., GOEL, V. & UPSHUR, R. 2004. Evidence-based health policy: context and utilisation. *Social Science & Medicine,* 58**,** 207-218.

DOYAL, L. 2000. Gender equity in health: debates and dilemmas. *Social Science & Medicine,* 51**,** 931-940.

EBELL, M., SIWEK, J., WEISS, B., WOOLF, S., SUSMAN, J., EWIGMAN, B. & BOWMAN, M. 2004. Strength of Recommendation Taxonomy (SORT): A patient-centred approach to grading evidence in the medical literature. *American Family Physician,* 69**,** 548-56.

EVIDENCE-BASED MEDICINE WORKING GROUP 1992. Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA,* 268**,** 2420-2425.

FIELDING, J. E. & BRISS, P. A. 2006. Promoting evidence-based public health policy: can we have better evidence and more action? *Health Affairs,* 25**,** 969-978.

GLASZIOU, P., VANDENBROUCKE, J. & CHALMERS, I. 2004. Assessing the quality of research. *BMJ: British Medical Journal,* 328**,** 39.

GOLDENBERG, M. J. 2006. On evidence and evidence-based medicine: lessons from the philosophy of science. *Social Science & Medicine,* 62**,** 2621-2632.

GRADE, W. G. 2013. *Grading the Quality of Evidence and the Strength of Recommendation* [Online]. Available: http://www.gradeworkinggroup.org/index.htm [Accessed 01/05/13.]

GREEN, L. W. & GLASGOW, R. E. 2006. *Assessing Generalizability (External Validity) of Evidence to Practice* [Online]. Hamilton: McMaster University. Available: http://www.nccmt.ca/registry/view/eng/157.html. [Accessed 12-06/13]

GREENHALGH, T. & WIERINGA, S. 2011. Is it time to drop the 'knowledge translation' metaphor? A critical literature review. *Journal of the Royal Society of Medicine,* 104**,** 501-509.

GROSS, S. 2009. Experts and 'knowledge that counts': a study into the world of brain cancer diagnosis. *Social Science & Medicine,* 69**,** 1819-1826.

GUPTA, G. R., PARKHURST, J. O., OGDEN, J. A., AGGLETON, P. & MAHAL, A. 2008. Structural approaches to HIV prevention. *The Lancet,* 372**,** 764-775.

HALL, J. 2006. *Microfinance Brief: Tap and Reposition Youth (TRY) Program*. New York: Population Council.

HAMMERSLEY, M. & ATKINSON, P. 1989. *Ethnography: Principles in Practice*, Routledge.

HAYNES, R. & GALE, S. 2000. Deprivation and poor health in rural areas: inequalities hidden by averages. *Health & place,* 6**,** 275.

HELMAN, C. & HELMAN, C. 2007. *Culture, Health And Illness.* London: Hodder Arnold

KOHLER, H.-P. & THORNTON, R. 2010. *Conditional cash transfers and HIV/AIDS prevention: unconditionally promising?* [Online]. University of Michigan. http://ipl.econ.duke.edu/bread/papers/working/283.pdf [Accessed 20/05/13]

LAVIS, J. N. 2006. Research, public policymaking, and knowledge-translation processes: Canadian efforts to build bridges. *Journal of Continuing Education in the Health Professions,* 26**,** 37-45.

LAVIS, J. N., ROBERTSON, D., WOODSIDE, J. M., MCLEOD, C. B. & ABELSON, J. 2003. How can research organizations more effectively transfer research knowledge to decision makers? *Milbank Quarterly,* 81**,** 221-248.

MCKINLAY, J. & MARCEAU, L. 2000. US public health and the 21st century: diabetes mellitus. *The Lancet,* 356**,** 757-761.

MEDLIN, C. & DE WALQUE, D. 2008. *Potential Applications of Conditional Cash Transfers for Prevention of Sexually Transmitted Infections and HIV in Sub-Saharan Africa.* Washington D.C.: The World Bank

MERTON, R. K. 1973. *The Sociology of Science: Theoretical and Empirical Investigations,* Chicago, University of Chicago Press.

MITTON, C., ADAIR, C. E., MCKENZIE, E., PATTEN, S. B. & PERRY, B. W. 2007. Knowledge transfer and exchange: review and synthesis of the literature. *Milbank Quarterly,* 85**,** 729-768.

MULROW, C. D. 1994. Rationale for systematic reviews. *BMJ: British Medical Journal,* 309**,** 597.

NICE. 2004. *Appendix A: Grading Scheme* [Online]. Available: http://publications.nice.org.uk/dental-recall-cg19/appendix-a-grading-scheme. [Accessed 20/06/13]

NICE. 2005. *NICE: Guideline Development Methods 11 - Creating Guideline Recommendations* [Online]. Available: http://www.nice.org.uk/niceMedia/pdf/GDM_Chapter11_0305.pdf. [Accessed 20/06/13]

NUGUS, P. & BRAITHWAITE, J. 2010. The dynamic interaction of quality and efficiency in the emergency department: Squaring the circle? *Social Science & Medicine,* 70**,** 511-517.

NUGUS, P., BRIDGES, J. & BRAITHWAITE, J. 2009. Selling patients. *BMJ,* 339.

NUGUS, P., CARROLL, K., HEWETT, D. G., SHORT, A., FORERO, R. & BRAITHWAITE, J. 2010. Integrated care in the emergency department: a complex adaptive systems perspective. *Social Science & Medicine,* 71**,** 1997-2004.

NUTLEY, S., POWELL, A. & DAVIES, H. 2012. *What Counts as Good Evidence?* [Online]. http://www.nesta.org.uk/library/documents/A4UEprovocationpaper2.pdf   [Accessed 20/06/13]

PAWSON, R., GREENHALGH, T., HARVEY, G. & WALSHE, K. 2005. Realist review - a new method of systematic review designed for complex policy interventions. *Journal of Health Services Research & Policy,* 10**,** 21-34.

PAWSON, R. & TILLEY, N. 1997. *Realistic Evaluation,* London, Sage Publications.

PETTICREW, M. & ROBERTS, H. 2003. Evidence, hierarchies, and typologies: horses for courses. *Journal of Epidemiology and Community Health,* 57**,** 527-529.

PETTIGREW, M. 2003. Evidence, Hierarchies, and Typologies: Horses for Courses. *Journal Epidemiology Community Health,* 57**,** 527-529.

PRONYK, P., HARGREAVES, J. R., KIM, J. C., MORISON, L. A., PHETLA, G., WATTS, C., BUSZA, J. & PORTER, J. D. 2006. Effect of a structural intervention for the prevention of intimate partner violence and HIV in rural South Africa: results of a cluster randomized trial. *The Lancet,* 368**,** 1973-1983.

ROTHWELL, P. M. 2005. Treating Individuals 1: External Validity of randomised controlled trials:"To whom do the results of this trial apply?". *Lancet,* 365**,** 82-93.

SILVERMAN, D. 2004. *Interpreting Qualitative Data: Methods for Analysing Talk, Text and Interaction,* London, Sage.

SILVERMAN, D. 2009. *Doing Qualitative Research*, London, SAGE Publications.

VICTORIA, C. G., HABICHT, J.-P. & BRYCE, J. 2004. Evidence-Based Public Health. *American Journal of Public Health,* 94**,** 400-405.

WILKINSON, R. G. 2002. *Unhealthy Societies: The Afflictions of Inequality*, London, Routledge.

WILKINSON, R. G. & MARMOT, M. G. 2003. *Social Determinants of Health: The Solid Facts*, World Health Organization.

YOUNG, T. K., SZATHMARY, E. J., EVERS, S. & WHEATLEY, B. 1990. Geographical distribution of diabetes among the native population of Canada: a national survey. *Social Science & Medicine,* 31**,** 129-139.