GETTING RESEARCH INTO HEALTH POLICY AND PRACTICE

GRIPHEALTH

Brief 4

October 2016

What is good evidence for policy?

Within many public policy arenas there are widespread calls to increase the use of evidence in policymaking. Often these calls rest on an assumption that increased evidence utilisation will be a more efficient or effective means of achieving programme goals. Yet, a clear elucidation of what can be considered 'good evidence' for policy is rarely articulated. Many current discussions on best practise derive from the evidence-based medicine (EBM) movement, embracing the 'hierarchy of evidence' that places experimental trials as pre-eminent in terms of methodological quality. This brief draws on insights from multiple disciplines to illustrate the limitations of a single hierarchy to guide policy decisions, and to construct a 'framework of appropriateness' through which to consider policy relevant evidence. In doing so, we are able to reconceptualise what might constitute 'good evidence for policy'.

The appropriateness of evidence for policy needs – moving beyond gold standards

EBM is seen to increase objectivity, transparency, and certainty of clinical practise. Since these are also goals espoused for policymaking, it is unsurprising that the logic of EBM – including the use of evidence hierarchies - has featured prominently in current discussions on the use of evidence in policymaking. Hierarchies of evidence typically place methodologies such as randomised controlled trials (RCTs) or meta-analyses of RCTs at the 'top', often referring to them as the 'gold standard' of evidence. However, such hierarchies were principally designed to judge evidence of intervention effect, not necessarily to reflect on policy relevance or importance(1) (see also Brief 2).

This, however, begs the question of what constitutes 'good evidence for policy' when policy usefulness is the principle criteria of concern. In this brief, we draw on the fields of policy studies, the sociology of knowledge, and philosophy of science to highlight challenges to the simple application of hierarchies, while further identifying alternative ways to evaluate what would constitute more 'appropriate' evidence to inform policy.

Policy studies: decisions involve multiple concerns

The first field useful to inform thinking on what constitutes good evidence for policy is that of policy studies. From this explicitly political perspective, two problems arise with the direct application of evidence hierarchies to guide policy decisions. First, policy decisions typically involve choice between competing concerns, not just technical evaluations of effectiveness; and second, those interventions conducive to experimentation may not be a policy priority. It is worth noting that these political realities do not eliminate the importance of evidence. Rather, to judge the extent

At a glance

- *Appropriate* evidence for policy, is that which speaks to the multiple concerns at stake, which is constructed in ways most useful to achieve policy goals, and which is applicable in the local context.
- *Good evidence for policy* can be defined as evidence which is appropriate (according to the above criteria) and which meets relevant quality standards from a scientific perspective.

to which a body of evidence is relevant, there will be a need to elucidate the goals and concerns of importance to the policy decision maker. Consequently, when judging whether evidence is 'good' we must assess whether it is relevant to the goals of the policy itself. Figure 1 provides a simple illustration of this, showing that of the entire field of evidence available, only a subset of evidence might be judged relevant to the concerns at hand. Some evidence outside of the subset will no doubt be of high quality, but this does not necessarily mean that it is evidence of usefulness to the policy decision at hand.

Figure 1

Thus, good practice in evidence-informed policymaking must start by making relevant policy concerns explicit. Without a clear indication of these it is impossible to say whether evidence is good or not, and it opens the door to so-called 'issue bias' in which the selection or promotion of particular pieces of evidence can obscure the relevant political concerns at hand (see also Brief 2).

Sociological perspectives: evidence is constructed (in more or less useful ways)

Sociologists have noted how social norms, ideologies, and power relations can be constructed into the creation of knowledge itself. This recognises that what counts as evidence is often an artefact of the context within which it is produced. Consequently, when reflecting on which evidence is most useful for policy purposes, it is important to recognise that there is often a choice of how to construct and classify data. For example, medical sociologists have explored how concepts like ethnicity or social class are often not adequately captured in health surveys or research, making it impossible for such factors to be the target of policy action(2). These insights allow us to question whether the categories and concepts used in a body of evidence are, in fact, the most useful to achieve policy goals.

Figure 2 thus illustrates how only a select range of evidence constructions will provide the most appropriate information for the policy goals at hand.

Figure 2

Philosophy of science: generalisability and evidence in context

Finally, work in the philosophy of science has specifically discussed questions around the generalisability of pieces of evidence. Specifically, when considering whether evidence is appropriate for a policy decision, it is necessary to consider if the results will apply in the local context. To do this, there is a need to consider the generalisability of any results – or, in more technical language, to distinguish between the internal and external validity of studies. RCTs are designed to have high internal validity – that is to show that they produced an effect where they were undertaken. They do not, however, say anything about the external validity – whether the result would be the same elsewhere. In other words, showing that an intervention worked in one place does not necessarily mean that the intervention works *always* and *everywhere* (c.f. Cartwright and Hardie, 2012)(3). This is especially true when dealing with social issues as interventions often work through alternative mechanisms in differing contexts. While biomedical interventions are assumed to be generalisable due to similarities in human physiology,

social interventions – such as providing cash transfers, or providing group-based health education – may work differently in different settings (or may produce an opposite effect if the context differs enough).

Figure 3, below, illustrates how bodies of evidence may be more or less relevant to the context addressed by the policy decision. Much evidence ranking highly on hierarchies in terms of rigour and internal validity, may not be applicable locally, and, as such, may not be appropriate for the given policy needs.

Figure 3

A framework of appropriateness

Taken together, these three perspectives provide clear reasons why hierarchies of evidence cannot serve as the sole measure of what constitutes 'good evidence for policy'. Yet, each perspective also provides insights about alternative ways to judge which evidence best serves policy needs. From them, we can define *appropriate* evidence for policy as consisting of evidence that addresses the political considerations at stake, that is constructed in ways that are useful to those considerations, and that is applicable to the local policy context. This can be illustrated by combining the small inner circles from the three previous figures, representing the subset of evidence that captures when these elements overlap (Figure 4).

Doesn't quality matter?

At this point, some may ask the question, 'doesn't quality still matter'? Of course evidentiary rigour and quality will always remain important. Yet, research can take many forms, and the way to judge quality will often be dependent on the type of evidence considered. As such, quality criteria should be decided only after there is an identification of which evidence is most useful to the policy concern. Specifically, the use of *methodological pluralism* is needed - an approach based on the principle of choosing the most suitable methods for the nature of the problem being researched and for which differing quality criteria will be relevant depending

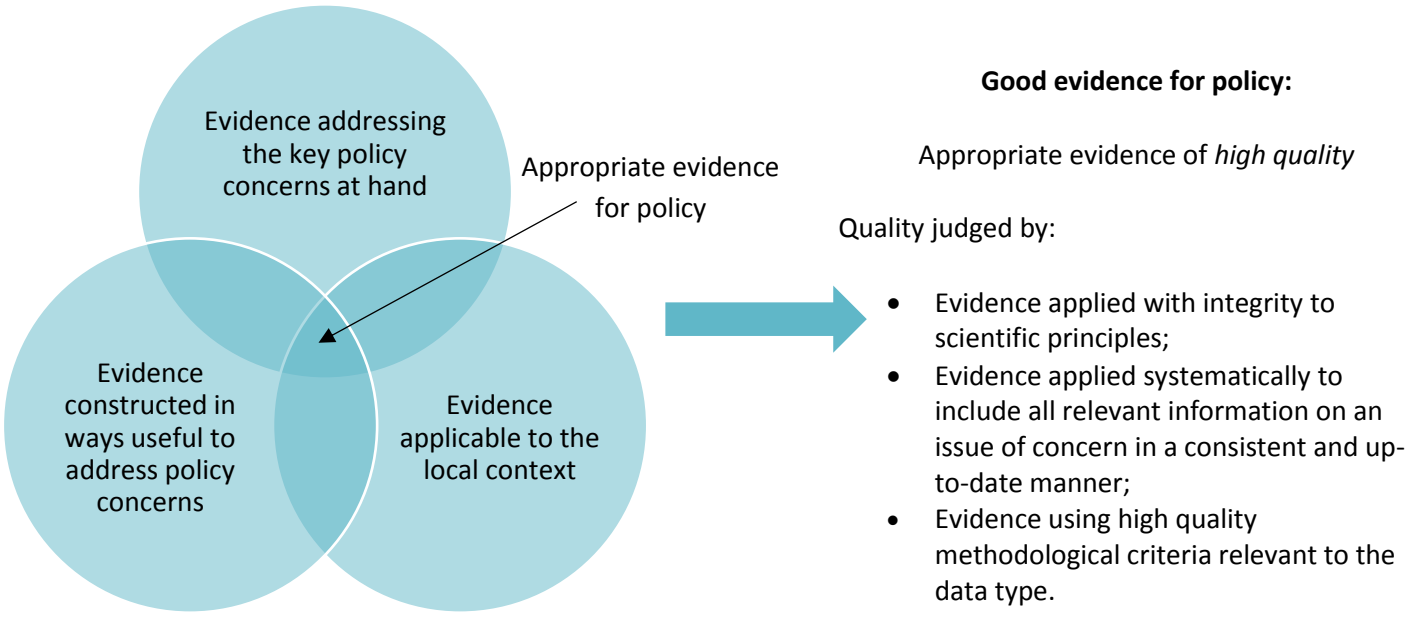

Figure 4

on the methodologies employed. So for example, if public acceptability is an important policy consideration, evidence from survey research may be appropriate, rather than RCTs – with survey quality judged by assessment of statistical power or representativeness. Thus, quality judgements can and should be used, but only after appropriate evidence is identified. Integrating the need for quality with the earlier discussion of evidence appropriateness, can thus allow a final definition of 'good evidence for policy' as *appropriate evidence of high quality*.

What about gaps in (policy-useful) evidence?

It is important to recognise that good evidence for policy, as defined here, does not equate to evidence of absolute certainty. In reality, many pieces of information of relevance to a decision may be unknown. Consequently, policy makers must often take action without complete information, and a judgement of when evidence is 'good enough' will be down to the individual decision maker. However, a lens of appropriateness can help to guide this decision, allowing for more direct reflection on whether the existing evidence base (including any gaps) is useful *enough* when the goals and needs of the policy decision are explicitly considered.

Discussion

Evidence use remains critical for improving and guiding policy decisions, yet there is a need to recognise the differences in needs and considerations at hand between public policymaking (including health policymaking) and the field of clinical medicine from which many EBP concepts have originated. In this brief, the question of what constitutes good evidence for policy has been reframed as a question of policy appropriateness, to move beyond over-simplistic applications of evidence hierarchies for questions they are not designed to address, and to help reconsider which evidence is most important to inform policy decisions. From this perspective, a set of strategic questions can be asked to guide reflections on evidence by decision makers:

1) Does the evidence address the multiple policy concerns at stake?

2) Are the data constructed in ways that best serve policy goals? And,

3) Is the evidence applicable in the local policy context?

Once these are considered to identify the appropriate evidence for policy, the evidence can further be judged in terms of its quality, however, based on the relevant methodological principles at hand for the type of evidence used.

References

1. Petticrew M, Roberts H. Evidence, hierarchies, and typologies: horses for courses. Journal of Epidemiology and Community Health. 2003;57(7):527-9.
2. Krieger N. The making of public health data: paradigms, politics, and policy. Journal of Public Health Policy. 1992;13(4):412-27.
3. Cartwright N, Hardie J. Evidence-based policy: A practical guide to doing it better. Oxford: Oxford University Press; 2012.

This brief is a summary of the chapter "What is 'good evidence for policy'? From hierarchies to appropriate evidence" in the book *The Politics of Evidence* available for free electronically from: http://bit.ly/2eQ3By2. An extended version also appears as Parkhurst and Abeysinghe (2016) "What constitutes 'good' evidence…" *Social Epistemology* http://bit.ly/2eiU4R5.

This brief was prepared by Siobhan Leir and Justin Parkhurst and is based on research conducted by the GRIP-Health programme, a research study supported by the European Research Council (project ID# 282118). For more information, please visit the GRIP-Health website at: http://blogs.lshtm.ac.uk/griphealth/