Name: Brendan Clarke, on behalf of the *Evaluating Evidence in Medicine* project team
Job title: Lecturer in History and Philosophy of Medicine
Organisation / institution: UCL
Is this input submitted as an organisational or individual response? Organisation
Are you happy for your response to be published by the Academy? Yes

---

1. The overarching aim of the workstream is to better understand how society uses evidence to judge the risks and benefits of medicinal products. In your view, what are the key factors underpinning this process that the Academy should consider?

There is currently vigorous debate about how evidence should be evaluated, within both the philosophy of science literature and the medical literature. We are an interdisciplinary research project composed of medical practitioners, philosophers of science, and other evidence stakeholders. We are currently working on a 3-year project research project (*Evaluating Evidence in Medicine: ebmplus.org*) that aims to investigate what we think is the key question underpinning medical decision-making: the role played by mechanistic evidence in medicine.

Evidence-based prioritises 'conscientious, explicit, and judicious use of current best evidence' (Sackett et al. 1996, *BMJ* 312:71). This 'best evidence' usually has a very specific meaning: the best evidence available to support decision making in medicine consists of statistical evidence that the putative cause and effect exhibit an appropriate form of correlation. This best evidence is produced by clinical trials, where treatments are tested on large numbers of patients. On the other hand, non-statistical evidence of mechanisms – produced, e.g., by experimental investigations in the laboratory or biomedical imaging – is held to be of low quality by most EBM practitioners.

However, recent work in philosophy has suggested that this hierarchy of evidence may be problematic (e.g., Cartwright & Hardie 2012, *Evidence-based policy*, OUP; Stegenga 2013, *Topoi* doi 10.1007/s11245-013-9189-4). In particular, evidence of mechanisms and evidence of correlation need to be treated more equitably (Clarke et al. 2014, *Topoi* doi 10.1007/s11245-013-9220-9.) This is for three main reasons. First, evidence of mechanisms is often crucial for setting up a clinical trial and for interpreting its results. Second, evidence of mechanisms is often crucial in order to extrapolate a causal claim from the population of patients in the trial to the population of individuals to be treated (on these two points see also Clarke et al. 2013, *Preventative Medicine* doi 10.1016/j.ypmed.2012.10.020). Third, establishing a causal claim normally requires establishing both that the putative cause and effect are correlated *and* that there is some underlying mechanism that can explain this correlation (Russo & Williamson 2007, *International Studies in Philosophy of Science* doi 10.1080/02698590701498084).

2. When evaluating the risks and benefits of medicinal products, what are the *strengths* of evidence that originates from different sources?

We make two claims in this respect. First, that evidence produced by different methods (trials, case reports, laboratory research) have different – but characteristic - strengths and weaknesses. These strengths and weaknesses can be characterised, and these characterisations can inform the use(s) to which evidence of different kinds can best be put. This is not a controversial claim when dealing with medical evidence – tying in closely with the current GRADE guidelines, for example.

Our second claim is that - because of the distinctive characteristics of different kinds of evidence - **certain combinations of evidence mutually strengthen and reinforce one another**. As we argued in a recent paper:

> *1. Evidence of a correlation relation between A and B:*
>
> *Its problems are that of confounding and non-causal correlations.*
>
> *Its advantage is that it can reveal masking, and can help assess the net effect of a complex mechanism.*
>
> *2. Evidence of a mechanism linking A and B:*
>
> *Its problem is masking, and being too complex to assess a net effect.*
>
> *Its advantage is that it can reveal confounding and non-causal correlations.*
>
> (Clarke et al. 2014, *Topoi* doi 10.1007/s11245-013-9220-9)

By way of explanation, we think that clinical trial evidence typically reveals correlations between an intervention, and a clinical outcome. The strength of this kind of evidence is that it reveals the overall, or net, effect of an intervention. However, it suffers from confounding, as is well known. On the other hand, we can also typically generate evidence of the causal mechanism that links our intervention and our outcome (often by laboratory research). On its own, this mechanistic evidence has characteristic strengths and weaknesses too. One strength is that this mechanistic evidence might be able to show that our clinical outcome arises causally, by virtue of the intervention, rather than coming from confounding factors. However, this kind of evidence suffers from a characteristic weaknesses, too: the mechanistic evidence typically will suffer from masking, which we define as follows:

> *…suppose you have very detailed evidence of a mechanism linking A and B. E.g., you have found the bacteria and understood how they cause the disease, you have studied an antibiotic and found that it kills the right bacteria, and doesn't harm people, and you are confident that killing the bacteria will cause full recovery. So you can trace the mechanism all the way from the antibiotic to recovery, or trace the process, in Steel's terms (Steel 2008). However, you cannot conclude that taking the antibiotic will cause recovery. This is because finding one mechanism linking A and B does not prove that there are no other mechanisms operating.*
>
> (Clarke et al. 2014, *Topoi* doi 10.1007/s11245-013-9220-9)

As can be seen from the above example, these kinds of evidence can be combined to mutually reinforce one another. For us, this combination of different forms of evidence is absolutely key to assessing the strengths (and weaknesses) of that evidence. We have likened this interaction between different forms of evidence to the way that the differing mechanical properties of concrete and steel are used to produce reinforced concrete:

> *We can describe this situation by an analogy to reinforced concrete, which is formed by placing steel grids into concrete. Now most concrete mixes have high resistance to compressive stresses, but any appreciable tension (e.g., due to bending) will break the microscopic rigid lattice, resulting in cracking and separation of the concrete. Steel, however, has high strength in tension. So, if steel is placed in concrete to produce reinforced concrete, we get a composite material where the concrete resists the compression and the steel resists the tension. The combination of two different materials produces a material that is much stronger than either of its components. In the same way,*

we argue that it is the combination of two different types of evidence which produces much stronger overall confirmation than would either type of evidence on its own.

(Clarke et al. 2014, *Topoi* doi 10.1007/s11245-013-9220-9)

3. When evaluating the risks and benefits of medicinal products, what are the *limitations* of evidence that originates from different sources?

As we discuss above, strengths and weaknesses of different kinds of evidence are intimately linked. Please therefore see answer to question 2 for a more detailed discussion.

4. Please provide details of any further examples or case studies that it would be useful for the project to consider.

Given the focus of this project, our most important suggestion here would be to include the consideration of medical examples that are not purely biomedical in nature. For example, thinking about the way that social evidence is gathered and interpreted would be most useful. One example here would be to think about the kinds of evidence that have been used when discussing the role of socioeconomic status [SES} in understanding disease causality (for instance, Freese and Lutfey 2011 doi 10.1007/978-1-4419-7261-3_4). Here, the assertion that SES makes a difference to many disease processes is not controversial. However, the way in which SES makes such a difference is extremely unclear. We think that one reason for the difficulties in determining how SES acts might be broadly understood as a consequence of the reluctance by medical researchers to engage with evidence that does not engage with molecular-level mechanisms of disease. We think this is regrettable.