

What Evidence in Evidence-Based Medicine?

JOHN WORRALL

Extended Outline

It is difficult to conceive of serious opposition to the general claim that the practice of medicine ought to be based on evidence. The devil, as usual, is in the details. The emergence and evolution of the Evidence-Based Medicine (EBM) movement raises a number of detailed methodological issues about evidence, and strength of evidence – issues of a kind that have, of course, been extensively studied by philosophers of science. I here concentrate on two sets of issues.

1. The role and import of randomisation

It was easy to get the impression from early accounts that the "E" in "EBM" stood exclusively for "evidence obtained from RCTs (randomised controlled trials)". ("If you find [a] study that was not randomized, we'd suggest that you stop reading it and go on to the next article" – Sackett (*et al*) (1997).) Later clarifications of the position, however, insisted that this was "misinterpretation number one" – and that EBM in fact allows that non-randomised trials are *sometimes* sufficient to establish therapeutic causation. Recent expositions of the view – and indeed, on closer reading, some of the earlier accounts too – seem to advocate the idea that, rather than being a black-or-white affair (no RCT, no objective evidence), evidential support for claims about therapeutic efficacy comes in varying strengths. An RCT provides the strongest, most "valid" evidence, non-randomised, but still controlled studies rather less strong evidence, case-controlled studies rather less still and so on. The EBM-er is then encouraged to base her decision on therapy on whatever is the strongest evidence available that is relevant to that therapy.

At least two problems arise. What underlying principles justify the evidential-strength ordering proposed? And, specifically, why exactly do RCTs carry most weight? This more specific problem of the alleged virtues of randomisation has already been studied by philosophers of science (see, for example, Urbach (1985)). Two arguments have drawn particular attention. The first is the argument, due to Fisher, that only if the division between experimental group and control group is made via some random process does the logic of the classical frequentist significance test really apply. The second is the rather looser one that randomisation controls for all possible confounding factors, known *and* unknown – at least "in some probabilistic sense". It is not clear that the first argument is sound and, even if it were, Bayesians would argue that the whole idea of classical significance testing is based on extremely shaky foundations. The second argument, which seems clearly to be the one that has persuaded the medical community that the RCT provides the "gold standard", is afflicted with a number of obscurities.

One interesting and independent argument which has often been cited holds that the chief rival to RCTs – historically controlled trials (ones in which the control group is provided by previous patients treated with conventional therapy) – are "known" to exaggerate the positive effects of proposed new treatments. The chief studies at issue (Chalmers *et al* (1977) and (1983)) looked at cases where RCTs and historically controlled trials had been performed on the same therapy and found a systematic tendency of historically controlled trials to favour the 'new' therapy compared to

RCTs. But the conclusion that historically controlled trials 'exaggerate' positive effects clearly follows from this finding only under the assumption that the true effect is accurately measured by the RCT. Moreover the finding itself has been challenged in some recent interesting papers (Benson and Hartz (2000) and Concato et al (2000)).

2. EBM as an "adjunct" to, rather than rival of, "traditional" approaches

As explained by Brian Haynes, "Phase One" of EBM tended to emphasise the "revolutionary" nature of the approach and the clash with traditional approaches (based on seeking the underlying physiological mechanisms of disease and therapy, and on "clinical judgment"). But "Phase Two" is altogether more ecumenical and cosy – seeing EBM as an "adjunct to" traditional approaches rather than a rival, and explicitly allowing an important role to "clinical expertise and judgment".

While "Phase One" EBM was undoubtedly too strong to be correct, the danger, of course, is that the ecumenical "Phase Two" may end up too weak to be interesting – EBM as "all things to all men" (as I have heard it described). In particular it is not clear what happens to the original motivating claim of EBM, one that accounted for much of its impact, that many treatments that medics' clinical judgments told them were effective, in fact appeared ineffective (or worse) when judged on the "objective" evidence.

The nature of the originally perceived clash with traditional approaches needs to be analysed. In particular, we need to investigate in what senses, if any, are clinical judgment and objective scientific method at odds. The current "Phase Two" account of evidence needs systematic rethinking from first principles. This account may seem to be precisely in line with some presently fashionable views in philosophy of science that insist that "disunity" – a sort of patchwork of methods with more or less *ad hoc* or vague ranges of application – is healthy and exactly the sort of thing we should expect on reflection in *any* science. However, I believe that a thorough rethinking of the role of evidence from first (and very general, unified) principles will explain this patchwork (or perhaps explain it with modifications – in the same way that Newton's theory explains Kepler's laws). The basic idea behind all blinding and control techniques is after all just the idea that one should not accept one explanation of some data, if other plausible alternative explanations exist (or, in other words, that theories should always be tested against plausible alternatives). When applied to differing particular circumstances, the intuitive idea that evidence for a claim arises from severe tests of it, and in particular severe tests of it *against plausible alternatives* goes a long way toward providing a unified and coherent account of the apparently disunified position adopted by sensitive EBM-ers.

REFERENCES

- Benson K. and Hartz. A.J. (2000) "A Comparison of Observational Studies and Randomized, Controlled Trials" *New England Journal of Medicine*, **342**, 25, pp.1878–1886
- Concato J., Shah M.P.H. and Horwitz R.I. (2000) "Randomized, Controlled Trials, Observational Studies, and the Hierarchy of Research Designs", *New England Journal of Medicine*, **342**, 25, pp.1887–1892

Chalmers T.C., Matta R.J., Smith H, Jr, and Kunzler, A.M. (1977)"Evidence Favouring the Use of Anticoagulants in the Hospital Phase of Acute Myocardial Infarction", *New England Journal of Medicine*, **297**, 1091–1097

Chalmers T.C., Celano P, Sacks H.S., Smith H Jr (1983) "Bias in treatment assignment in controlled clinical trials", *New England Journal of Medicine*, **309**, 1358–1361.

Sackett D.L., Richardson W.S., Rosenberg W., and Haynes R.B.(1977)*Evidence-Based Medicine: How to Practice and Teach EBM*, New York: Churchill-Livingstone.

Urbach, P.M. (1985) "Randomization and the Design of Experiments", *Philosophy of Science* , **52**, pp.256–273