IN THIS ISSUE…

Choosing Treatments for Patients: The Not so Simple Case of Basal Cell Carcinoma

Anybody interested in asking how we should treat basal cell carcinoma (BCC) should read the review paper from Mosterd et al. in this issue of the journal (p. 454–458). This article, from Kelleners-Smeet's Maastricht group, provides for some readers, if not for the authors, a fascinating case history of the deficiencies of the evidence-based medicine (EBM) movement (1). Let me first describe the paper briefly and then explain its wider significance.

Mosterd and colleagues have reviewed the randomized controlled trials (RCTs) of treatments for BCC published in the Cochrane Reviews. They list the RCTs, cluster them based on known pathophysiological groupings of BCC, and attempt to determine the optimum treatment amongst the various modalities available (simple excision, Mohs' surgery, photodynamic therapy, etc.). Few will challenge the fact that they have achieved the first two of these goals (listing and clustering), but the interest is obviously in the third: which treatment for which patient? In a wonderfully honest understatement the authors rightly point out that, "many problems remain unsolved concerning the treatment of BCC", and they attribute this to the dearth of RCTs and problems with quantification of clinical outcomes. They are right (in part), but some of their arguments deserve a more expansive and coherent approach.

If we are forced to choose between evidence based on a RCT on the one hand, and on simple observational clinical practice on the other, we would always chose that from the RCT *if all other factors are kept constant*. A RCT allows us not to have to worry that the allocation of treatments to patients is biased: trial randomization acts so as to diminish any intentional or unintentional malicious linkage between patient characteristics and treatment allocation (at least in the long run). However, the devil is in the phrase "if all other factors were kept constant". We know that RCTs form a small subset of medical knowledge, especially in dermatology. Most of our knowledge comes from non-randomized observations. This can be for at least three reasons. First, as Mosterd et al. point out, there are simply very few RCTs in many clinical areas. Secondly, much of the evidence needed to guide clinical practice cannot be studied in RCTs: long-term rare side-effects would be one example (think of neurological viral disease or cancer after biologics), but there are many others. Thirdly, and now the subject of a large volume of literature, the patients in trials are not a random subset of patients with a particular disease. The importance of randomness here is not dissimilar to the reason why we prefer randomization once patients are chosen to take part in the study – it allows a sound statistical basis for inference from sample to larger population. Let me explain this in more detail.

The beauty of randomization between treatments once the subjects have been chosen for a trial is that it allows us to infer treatment efficacy for a notional large population of subjects if the study were to be repeated many times. It allows us to have confidence in what would happen to future patients if they had the characteristics of those subjects taking part in our current study. The difficulty arises that we are not only interested in the sorts of subjects who participate in our studies, but rather in the totality of patients who we see in clinical practice. How do we infer what works for them? Well, in a word, judgement. We have to leave behind our statistical summary measures because they refer to trial populations, and instead we have to judge how similar patients in our own clinical practice are to those who took part in the study. However, this is not a simple statistical judgement (you cannot offload this to your statistician, nor can he or she produce confidence limits to describe this uncertainty), but instead a clinical judgement that has to take the totality of evidence into account. So, what is the totality of evidence, and where does the exclusion of all the non-RCT evidence lead?

The principal aim of the EBM movement has been to demarcate RCT evidence from non-RCT evidence (1). As stated above, if all other factors were constant, this would be reasonable. But we know all other factors are not constant, and excluding evidence is anathema to any serious scientist. Imperfect experiments are still recorded because all experiments ultimately are imperfect. Despite a plethora of recommendations from committee after committee we still do not have any formal way to summarize randomized and non-randomized data into measures of effect and certainty. Of course, bureaucracies love grading evidence – it fits nicely into Excel spreadsheets and (pun intended) it is an evidence-free zone. However, the core philosophical problem is that the certainty with which one holds a belief does not have a simple relation with how the evidence was obtained (2). This has troubled logicians for a long time and its wider promulgation would act as a wonderful disinfectant for the epidemic of verbiage some of us are forced to endure. I can be more certain that retinoic acid causes foetal malformations than that one antifungal studied in a handful of trials is better for my patients than another. Imagine two six-chamber revolvers, one with six empty chambers, the other with only one of six empty. If forced, which would you prefer to use in your defence. Well, don't look to a Cochrane Review for help. I may trust an observational study more than a RCT if the patients who took part in the RCT do not resemble the patients I see on a daily basis. What I will not be able to do is attach any confidence limits based on classical statistics to this judgement. Is there any sensible way out of this impasse? Perhaps.

The essential goal of treatment summaries, meta-analyses or conventional reviews such as those found in textbooks, is to guide treatment. For the reasons outlined above there is no absolute demarcation between RCT and non-RCT evidence; there is no ranking possible between EBM textbooks and review articles, and between conventional textbooks such as those of Braun Falco or Rook. Instead, we must view books or guidelines as a set of instructions that guide practice, much as we view computer code as a set of instructions that allow a universal computer to produce an output in response to a given input. The analogy with computer code is, I believe, informative. Maurice Wilkes, one of the fathers of modern computing, described in a flash of insight how, in the late 1940s, he realized that he would spend most of his life debugging the programs he had already written (3). Sets of instruc-

tions can be dashed off quickly, but for all but the simplest code, it takes longer to debug than write the original. Despite all the brains and input – far greater than usually take part in committees summarizing evidence – most computer programs remain full of poorly-defined bugs. We accept that the code does not just need thinking about – very clever people have already looked at it – but rather we have to implement it and measure the outcomes. The software is released as a "beta" version, large numbers of users try it, and there is an iterative feedback loop to the creators, and the cycle continues. By contrast, the missing step in most reviews of treatment efficacy or committee-produced guidelines is that they resemble ex-cathedra statements, hovering between the trivial and the incommensurable, with little or no evidence that when they are implemented they produce the desired output. This is a startling state of affairs. The irony, for this essay at least, is that it may matter less how the treatment decisions are arrived at – personally I would prefer Braun Falco to any committee guideline – than an empirical test of what happens when the recommendations are followed. What we require is a form of meta-knowledge; empirical knowledge about knowledge. Ironically, we either need observational evidence to assess RCT evidence, or we need to think of ways to test competing guidelines experimentally, perhaps using some form of simulation technique. Until then I suggest you look at Mosterd and Braun Falco and discuss them with your colleagues.

REFERENCES

1. Rees JL. Evidence-based medicine: the epistemology that isn't. J Am Acad Dermatol 2000; 43: 727.
2. Haack S. Puzzling out science. In: Manifesto of a passionate moderate. Chicago: University of Chicago Press; 1998, p. 90–103.
3. Wilkes MV. Memoirs of a computer scientist. Cambridge, MA: MIT Press; 1985.

Jonathan Rees
Section Editor