Reviewer's report

Title: Making trials matter: pragmatic and explanatory trials and the problem of applicability

Version: 1 Date: 18 February 2009

Reviewer: David Kent

Reviewer's report:

The authors present a narrative review contrasting pragmatic and explanatory trials, and weigh in clearly in favor of the pragmatic attitude (while conceding that explanatory trials have an important role). The authors are clearly thoroughly familiar with the literature, having made important recent contributions, notably as part of the CONSORT group. Overall, I found the paper enjoyably thought provoking—but not least because I was not wholly persuaded by their main argument that trials generally should be more “pragmatic” (although I would agree that “pragmatic” studies should more often be done after efficacy has been established, with some concerns and caveats).

My first 3 comments reflect counter-arguments that do not seem especially well-defended against in the manuscript. While these are my major concerns (and listed as “major compulsory revisions”), and I want to be sure that these points have been fully considered, the authors may chose not to substantially revise their manuscript.

Major compulsory revisions:

The author’s main argument is that many trials that are done should be more “pragmatic” and less “explanatory”. This argument would be considerably strengthened if it were buttressed by actual examples where pragmatic (or effectiveness) trials would have likely been more informative than explanatory (or efficacy) trials, or when substantial harm resulted because of efficacy trials. While there are sure to be examples of this, when I tried this thought experiment on actual trials, it became clear that efficacy trials are typically the way to go, especially for clinical trials trying to advance the technological/therapeutic frontier—even if they are often “aspirational” in many settings. The reason for this is that efficacy trials are very informative when they are either positive or negative, while effectiveness trials may be informative when they are positive and less so when they are negative (particularly if ideal care is compromised in many dimensions simultaneously). The FDA guidance is clear on this point; a trial may be negative for many reasons. The therapy may just not work. Or the therapy may work in patients that were more carefully diagnosed, had less competing risks, had better compliance, or with better monitoring, or in settings with a higher commitment/more experience with the intervention. (Or the therapy
might even work in the enrolled patients but the treatment signal was obscured by measurement error, etc.). So, your pragmatic trial is null, now what? Indeed, uninformative negative pragmatic trials may be quite common, given all these potential influences toward the null. Can we think of an example of a positive (or negative) efficacy trial that would have been more informative if done as a pragmatic trial with null results? On the other hand, an efficacy trial which shows null results, or even minimal benefits, even under the best of all possible circumstances, informs us better about low value therapy.

Advocates of pragmatic trials typically neglect the importance of trials being “aspirational”. Many new technologies/therapies do require structural or organizational changes to be effectively implemented. Testing them initially in usual care settings without substantial investment in organizational changes would be uninformative. The NINDS trial showed that thrombolytic therapy can be effective in well selected patients in settings highly committed to the treatment of acute stroke. It also showed that there are substantial potential risks to this therapy, consistent with prior evidence. Many practitioners appropriately felt that the results did not generalize to their setting, and appropriately resisted rapid adoption of thrombolysis. However, substantial efforts were devoted to implementing quality improvement programs, and certification processes, so that hospitals could achieve low rates of protocol violations, reasonable door-to-needle times and acceptable complications rates. There are many similar examples (surgical and other procedures, such as CABG, CEA, PCI, etc). While further effectiveness studies are necessary, such studies should be used to help find a feasible path forward. In this context, replacing efficacy studies with more “pragmatic” effectiveness studies seems like a recipe/excuse for inertia.

A related problem that advocates of pragmatic trials frequently gloss over is the idea of generalizability. The concept of generalizability is central to the argument for more pragmatic trials. The implication is that if one includes all different sorts of patients, in all different sorts of settings, somehow means that the overall results generalize to all included patients, in all included settings. This does not make any sense to me. Assuming a single treatment effect applies to everyone in the trial makes sense only in the absence of substantial treatment effect heterogeneity. However, the authors seem to take as a given that treatment effects vary substantially in different patients and in different settings—and justifiably. If this is the case, including older patients, patients with multiple co-morbidities, patients with different disease severities and patients from all different settings will surely influence the overall average but it does not mean that this overall result applies to all (or even most) included patients (see e.g. Rothwell, Lancet and Kent/Hayward JAMA, 2007). While subgroup analyses may make these trials more informative, the limitations of subgroup analyses are well-known. So-called pragmatic trials, with extreme patient heterogeneity, may yield results that are uninterpretable (and not clinically applicable for doctors who treat individuals in specific settings and not populations across settings), unless accompanied by especially well-thought out and rigorous subgroup analysis. This point often seems to be under-appreciated by advocates of pragmatic trials, and is not addressed in the current manuscript.
Minor compulsory revisions:

The authors cite numerous recent review articles on the same topic, including several due to be published shortly in JCE. It is not clear what is new and different about this review. While the authors hope that the review will be of use to trialists designing clinical trials and cite several useful tools, these tools are not presented in sufficient detail to be useful. The authors should address what this narrative review adds to the many recent articles they cite (although this need not be done directly in the manuscript).

At times, I found myself a bit confused as to the exact point the authors were trying to make. In particular, the issues covered on pages 6/7 were a little lost on me. I am hoping that the authors can clarify this section.

Confusing sentence, page 7 (sic): “of 26 trials, 13 of 18 were judged to be effectiveness trials...”