Author's response to reviews

Title: Association of Trial Registration with the Results and Conclusions of Published Trials of New Oncology Drugs

Authors:

Nicolas Rasmussen (N.Rasmussen@unsw.edu.au)
Kirby Lee (leek@pharmacy.ucsf.edu)
Lisa Bero (berol@pharmacy.ucsf.edu)

Version: 2 Date: 9 December 2009

Author's response to reviews: see over
Dear Dr Vickers and TRIALS Editorial Team

Please accept the revised manuscript accompanying this letter, with the point by point response to our referees describing how our manuscript now reflects the changes and clarifications they sought (below).

As you will see below, although the attached version still includes a discussion of sample size and power, we leave to your discretion the choice of whether this should be omitted. If so, please let me know and I will quickly make the deletion along with a few small changes to the paragraph involved.

Best

Nicolas Rasmussen

(With Lisa Bero and Kirby Lee)

Nicolas Rasmussen, MPhil, PhD, MPH
Associate Professor
School of History & Philosophy
Morven Brown 314
University of NSW
Sydney 2052 AUSTRALIA

RESPONSE TO COMMENTS, MS 7956236643146529.

Comments from Editor

First, the effect size sought by the authors is, in my view, misplaced. Registration is highly unlikely to reduce the percentage of positive trials from 75% to 50%; this would imply unprecedented levels of publication bias. My guess is that we demand registration because it may help bring to light a few trials per 100 than may not otherwise be published. The results of these trials might have implications for the medical care of thousands or even millions of patients. So even if registration (which is a trivial expense compared to the cost of a trial) had a very small impact on the literature, it would be worth doing.
Second, because of the limited sample size, the authors’ multivariable model is unable to adjust appropriately for confounders. …. In brief, the results have to be way toned down: this study is just too small to generate conclusions about the value or otherwise of registration.

Some minor points: please report median and quartiles not median and range; give year of publication in categories not means and SDs (from which we could calculate that 2.5% of papers in the prior registration group were published in 2010); use log transformed or square root of sample size in the multivariable analysis, not quartiles.

RESPONSE

First, the editor’s expectations of a very modest effect of registration on publication of favorable outcomes is pessimistic, albeit plausible. Unfortunately that level of pessimism is rare, as evidenced by the commonplace expectation that the policy of requiring registration will be adequate to control bias in the RCT literature. Testing this expectation in the central purpose of this paper. The effect size implied by the power calculation is roughly the effect size one would want to see if registration really was a powerful solution in itself, given the odds ratios over 4.0 typically found in both in oncology literature studies and in broader meta-analyses comparing reported efficacy outcomes in publicly vs pharma-sponsored RCTs (which we cite in our introduction). Please see further comments on equipoise below in reply to Dr Djulbegovic, points 1-3.

However, we would be willing to drop the entire discussion of sample size and power (Results, first para, first 2 sentences) on the grounds that these calculations require judgments. The relevant first paragraph of Results remains unchanged in the revised draft as resubmitted, but we will accept the editor's decision on whether to delete the first 2 sentences and thus the topic of power.

Second, the body of oncology literature we analyze has many epistemological advantages for addressing the impact of registration, although the sample size is inherently limited (as the editor seems to appreciate). In accordance with the editor’s suggestions we have now carefully edited the entire paper to "tone down" our conclusions and avoid any suggestion that trial registration is not useful. We have not deleted any results from our paper as we think this would introduce a reporting bias. We state in both our abstract and discussion that we are not suggesting there is no benefit in registration. Our findings suggest that registration should be expanded to be more useful, specifically to include full reporting of results and protocols. *

Finally, we have adjusted the data analysis and presentation in accordance with the editor’s suggestions. In particular: (1) Table 1 now reports Year of Publication in quartiles; also, Impact Factor and Sample Size are both expressed as median and interquartile range. (2) Tables 3 and 4 (the multivariate) now treat Sample Size as a single numeric predictor, lnX. This latter change to the model does not materially affect our conclusions as to Registration and to Blinding and (surrogate) Outcome Measure as predictors of Results. It does however elevate Sample Size to significance as a predictor
of Conclusions. Since (favorable) Results is our primary outcome measure we do not elaborate on this new finding.

[* I might note that the multivariate analysis described in the first version of this paper employed the Hosmer-Lemeshow progressive model building approach, mitigating the ‘too many variable’ issue to some extent. -- NR]

Reviewer: Benjamin Djulbegovic

Reviewer's report:

This is a well written and important paper. I have no major feedback/critique to offer. My comments are related to my own thinking about the issues the authors described in the paper that may be of potential usefulness to them should they decide to incorporate some of it in the paper.

General comments
1. the key issue in this type of research revolves around hypothesis about the successes of new, experimental vs. standard treatments. Any conclusions are then drawn by comparing observed with hypothesized results. The authors write “We hypothesized that advance trial registration is associated with statistically non-significant efficacy outcomes in published trial reports, when controlling for other known predictors of reported results such as funding source, publication characteristics, and trial design characteristics.” While I wholeheartedly agree with this hypothesis, the authors should really say why they hypothesized that success of new treatments should, on average, be about equal to successes of standard treatments. To the best of my knowledge, the only rationale for this hypothesis could be found in so called “equipoise hypothesis” [ref #16; see also Cancer Control 2009;16:342-347; J Med Philosophy. 2007;32:79-98; Arch Intern Med. Mar 24 2008;168(6):632-642] (Sorry for citing my papers, but I am not aware of others that explicitly postulated the hypothesis explaining the pattern of treatment successes observed in clinical trials).

RESPONSE: Our null hypothesis and power calculation were influenced by Dr. Djulbegovic’s important work establishing that the high quality Phase III trials run by the NCI Cooperative Oncology Groups have in the long run favored old treatments very nearly as often as new treatments, and in that sense conform roughly to the equipoise premise. This background represents another important advantage of studying registration’s effects in oncology. We had cut a short statement to this effect from an earlier draft in the interest of space, and in the attached we have restored it to the introduction (third para), with a citation to Dr. Djulbegovic’s key 2008 paper.

2. If we accept the “equipoise hypothesis”, then the question becomes are the
results still consistent with it? Although quantitative data still indicates this to be
case, the qualitative data seem to significantly favor new drugs (80% of time in
the pre-registration cohort of the trials). The authors (and I am inclined to agree)
believe that the result is the consequence of bias in reporting. However, industry
folks claim that they are so careful in designing trials that their success rates is
much higher than predicted by the equipoise hypothesis. So, who is correct? One
reason that I think the observed results are due to biased reporting- notably due
to publication bias- is that the authors only studied drugs approved by the FDA.
However, drugs that were never approved are neither registered nor published
(the FDA refuses to release these data due to “proprietary reasons”). I think this
is important to acknowledge.

RESPONSE: We did not add a comment on the industry defense because there is no
evidence that industry trials are more rigorous methodologically and some studies that
have examined reporting bias have controlled for methodological rigor and still observed
the bias. We have added as a limitation that we do not examine drugs that were not
approved by the FDA as we are not able to obtain this list (Discussion, 4th para). We did
d of course include all trials of the new oncology drugs in the sampling interval, not just
trials for the approved indications.

3. I am curious to learn how the authors calculated the “effect size” i.e. that
unregistered trials will be associated with 75% of success rate. This is very
important assumption.

RESPONSE: Given first, that Dr. Djulbegovic has given us good reason to expect that
the underlying odds of favoring new drugs among high quality trials in oncology (both
published and unpublished) are close to equal, and second, that there is evidence in the
literature suggesting publication bias and other forms bias accounting for odds ratios
upward of 4.0 when comparing efficacy outcomes in published pharma-sponsored vs
unsponsored drug RCTs, the premises of our power calculation are quite reasonable. But
as noted above, we are willing to drop the entire discussion of power.

Minor issue (points that need clarifications)
- how did authors handle multi-arm trials? Which treatment was selected as new
and which as standard RX?

RESPONSE: These questions are addressed in the Methods section of the manuscript
(subsection Comparison Group): in multi-arm trials the placebo arm was coded as
comparator in preference to active therapy (so as not to ‘disadvantage’ test drugs because
trialists tested both); and the ‘new’ or ‘test’ drug was defined by its month of FDA
approval. And as noted in the Results, this approach produced only one instance in
which a trial was sponsored by the maker of a drug other than our ‘test’ drug so defined.

- do you have data on accelerated vs. regular approval? This analysis may be
interesting.

RESPONSE: We did not do subgroup analysis by approval type because our numbers were limited.

- what was percentage of phase 2 RCTs vs. phase 3 RCTs? Any impact on results?

RESPONSE: Again, we did not conduct this subgroup analysis due to our limited sample size. In addition, this would have been problematic as quite a few trials that were conducted outside the United States did not apply this distinction in the FDA manner and besides, a large number of trials were self-described as “Phase 2/3”.

- empirical evidence shows that blinding is really not important when it comes to “hard” outcomes such as survival. So, not sure that was important to have such elaborate coding for blinding.

RESPONSE: Blinding has been shown in numerous studies to be associated with reported outcomes so we considered it important to include in the analysis. Due to the relative lack of importance of blinding for outcomes such as survival, we suspected that trials with survival as the primary outcome measure would tend to be less stringently blinded. However, interaction testing showed no such effect. A sentence stating “There was no interaction between stringent blinding and surrogate outcome measure, or among any other variables” has been inserted (Results, 3rd para) for emphasis.

- although it goes without saying that survival is the most important outcomes, in most of these trials progression-free survival (PFS) was primary outcome. Will result change if PFS is selected as the most important outcome?

RESPONSE: We did not address this question, but it probably would not have been feasible. Not all trials reported PFS as an outcome measure. Many reported similar but nonequivalent measures such as response rate.

Reviewer: Joel Lexchin

Reviewer's report:
This paper compares clinical trials that were registered with those that were not to see if registration affects biases in the reporting of results. In general I found this to be a well written paper that employed appropriate methodology. I have relatively few comments to offer.

Minor essential revisions
Page 9, first paragraph:
How was "most emphasized" operationalized, e.g., by a word count?

RESPONSE: First, we selected the outcome measure discussed at greatest length in the results section of the papers, unless it was not described in both the abstract and conclusion. In the rare cases that this occurred, we selected an alternative outcome by word count in the abstract. All disagreements were resolved by discussion between two authors. We note that this item was double-coded with high agreement (though kappa cannot be calculated because there are an indefinite number of potential main outcomes among those studies not explicitly designating primary outcome measures).

We have added the two clarifying sentences to this effect in the Methods section of our paper (at end of subsection “Type of primary outcome measures”)

Page 14, second paragraph:
The authors should also note that they only looked for differences in reported efficacy and not in reported safety issues.

RESPONSE: We have added this clarification to the existing sentence (Discussion, last para): "We examined only the primary efficacy outcomes of the trials, and thus cannot determine whether trial registration might influence the reporting of secondary outcomes or safety outcomes."

Page 14, Conclusions:
I would like to see a more fulsome discussion of new policy options that flow from the authors' findings.

RESPONSE: We have clarified the final paragraph of the paper (Conclusions section). The main policy options we propose are registration of full results and the full registration of protocols. We have added another sentence suggesting what might be done if design-based bias remains.