Review of: "How many papers are published each week reporting on trials of interventions involving behavioural aspects of health?"

Robin Delabays[1]

1 HES-SO Valais-Wallis

Potential competing interests: No potential competing interests to declare.

Based on bibliometric data, the authors provide an estimate of the number of papers published every year studying behavioral aspects of health in a broad sense. This estimate is then used to assess the amount of resources that would be needed to analyze the newly published literature every year. The authors' conclusion is that automating the extraction of information from publications would spare a significant amount of resources and they suggest to work towards a standardized machine-readable publication format.

As a preamble, I would like to make clear that I agree with the authors conclusion regarding formatting of research outcome. In my report I will however play the devil's advocate, pointing out to places where, in my opinion, the message could made more powerful.

The paper is well written and the message is very clear. I think, however, that the bibliometric analysis could be made more thorough with a moderate amount of work.

First of all, the overall message that the amount of scientific publication is large and increasing is quite well-documented in the bibliometric literature. I think that a bit more of context would provide a clearer overview of the state of the art in the domain. As a starter I would suggest [Price (1965)] and [Bornmann and Mutz (2015)].

Another point that would strengthen the analysis would be to provide the estimates with margins of error. Indeed, so far, there are estimates (e.g., 106 papers per week), which are acknowledged to be uncertain, but there is no quantification of how uncertain they are. There is also no details on how the estimation of the resources needed (in terms of person-hour or in terms of money) are done. A bit more detail would be much more convincing for the reader.

Finally, it is clear that at least a part of the work of extracting pertinent information from the literature is already done nowadays. This work is actually one of the jobs of any scholar. Providing an estimate of the current amount of this work that is already done would give an idea of the amount of resources that could be spared by an automation of the process.

I also add a few minor comments that will be easy to fix and would, in my opinion, improve the paper:
- In the list of reasons for the possible inaccuracies at the end of the introduction, an example for bullet point nr. 4 would be appreciated;
- In the first paragraph of the Methods section, it is unclear if the authors consider PubMed to be "comprehensive" (first

sentence) or not (last sentence);

- The expression used in the database is a bit hard to parse. I would suggest to list the terms in a table;

- Maybe the authors could explain why they restricted their search to years after 2018. Why not 2016, or 1998, or as late as possible in PubMeb?