Firstly, the topic of this systematic review is an important and useful one in the field of health service research. It indicates that instruments for measuring research use in the healthcare context have been significantly underdeveloped.

In general, the background section is very informative and coherent. The information gap between what we already know, and what we need to know, and why, is well described based on the chosen theories (Weiss, Dunn and Rich) and the empirical evidence (Estabrooks and colleagues). The need for this systematic review is clearly justified.

The purpose of the review is to assess the psychometric properties of instruments to measure research utilization in healthcare. Specific objectives are both to indentify instruments used to measure research utilization by healthcare providers, healthcare decision-makers, and in healthcare organizations and assess the psychometric properties (reliability and validity) of these instruments by using the Standards for Educational and Psychological Testing. However, I cannot find the answers for all these specific questions from the results section. In addition, I am wondering, why you do not use the Standards for Educational and Psychological Testing (pages 25-36) for assessing the reliability aspect in the same way as validity?

The method section describes well, what you have done and how. The study selection criteria, the search strategy for identification of studies, the selection of studies, the data extraction and the data synthesis are reported carefully. Also the figures and the additional files give more precise information for those with a deeper interest.

The discussion is organized around three main areas, which are composed of several specific sections. Reasoning is well explained, and you have also given
some explicit recommendations or caveats for researchers (e.g. pages 20-21).

Major revisions:

First, you have reported characteristics of the research utilization instruments in the results section (page 12): “Most studies used the instruments with clinical care providers…” However, I ask that you answer the specific study questions precisely both in the abstract and the results section, for example, how many instruments you identified among healthcare decision-makers. Furthermore, who are those healthcare decision-makers to whom you refer? Are they political decision-makers, leaders of healthcare organizations or decision-makers in patient treatment?

Second, you have reported the reliability of the identified research utilization instruments in the Reliability section of Psychometric Assessment of the Research Utilization Instruments (pages 13-14) and in the table 5. But, I feel that it is not in balance with more deeply analyzed validity metric according to the Standards, specifically, the reliability section in the pages 25-36. Because, reliability is mentioned in the title of your systematic review, you should review this apparent imbalance in the discussion section.

Third, the terms ‘acceptability’ and ‘feasibility’ (as psychometric properties) are only mentioned for the first time in the results section (page 13). I ask you to consider, how you can better engage these concepts in this review, and what we already know about these? I notice that you have mentioned lack of assessment of acceptability and feasibility as additional limitations in the Field (page 22). What do you mean by these concepts?

Fourth, you have defined research utilization in two ways, as 1) “the process by which specific research-based knowledge (science) is implemented in practice” according to Estabrooks and colleagues (page 4), and 2) the use of research-based (empirically derived) information (page 8). The last one is more precisely defined than the first one. According to the second definition, you decided to exclude articles that reported on healthcare providers’ adherence to clinical practice guidelines. One possible rationale for this decision is that clinical practice guidelines can be based on non-research evidence, e.g. expert opinion. That is true, however, I think that although your exclusion is well justified, you should discuss this decision in the discussion section. Particularly, because your data includes only seven studies targeted to physicians (page 12).

[Furthermore, I see that your first definition includes also clinical practice guidelines, because they are internationally used way to implement research-based knowledge in healthcare practice and providers. According to Varonen and Mäkelä (1997) 83% of the national guidelines were supported by a systematic literature analysis. Useful criteria have been developed for better guidelines (Grol et al. Improving Patient Care, 2005, pages 71-92) and classification of the literature (GRADE) according to the strength of the evidence as levels A, B, C, D.]

Minor revisions:
1. I hope that you reconsider the title of the review, Measuring Research Use in Healthcare Organizations, because this healthcare organization view you present is limited and you do not describe clearly what you mean by it.

2. In the abstract, you report that “we also hand searched specialized journals…” in plural although it may need to be in the singular, because in the search strategy for identification of studies section you report “we also hand searched the journal Implementation Science (a specialized journal in the research utilization field) in the singular.

3. In the last paragraph of the background section, you argue your reasons for why you believe that the measurement of research utilization is important. You mention a “black box” of causal mechanisms that influence research utilization without any reference. It would be useful if you could attach the word can to this sentence or report some references.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests.