Review of: "Kalya Research: Complementary and Alternative Medicine (CAM) Virtual Research Assistant from Biomedical Literature"

Kaitlyn Hair[1]

1 University of Edinburgh

This is an interesting manuscript describing the development and evaluation of a new tool to help identify relevant literature in the CAM domain. I have several queries for the authors and suggestions for improvements prior to publication:

1. The methodology for the selection of relevant articles is described, but the methodology underlying the named entity recognition and study classification is lacking in detail. For example:

   1. How did you determine "countries interested in these research issues"? Is this extracted from the first author's affiliation?
   2. How are you determining diseases, outcomes, number of participants, and health goals? Have you created software for this, or are you using pre-existing tools or existing meta-data?
   3. How did you determine the study type (e.g., review, trial) - is this information coming directly from PubMed?
   4. Which sources were used to collect articles from? You mention PubMed and "other journals specialized in CAM but not referenced in PubMed," but a comprehensive list would be helpful here.
   5. Are you using full-text articles or the full text of publications for NER?

2. I couldn't find any mention of when the PubMed search(es) were conducted or how up-to-date the database is. Does KR continually bring in new articles?

3. The tool aims to support researchers to find relevant publications, but there is little mention of the strength of the evidence and/or study quality (e.g., measures to reduce bias, sample size). At present, articles are ranked based on citation metrics and other criteria. Are there any plans to assess study quality in an automated way to support evidence-based research evaluation?

4. The comparison of searches in Medline vs. KR is a useful addition to demonstrate the utility of the tool versus traditional approaches. This could also indicate the % of included articles (from the initial search number). In addition, can you reflect on why (from a technical perspective) KR performs better? You mention that you filled the search bar in KR with 'breast' and the outcome field with 'alopecia' - does this indicate that the intervention part of the Medline search was too restrictive versus your NPI rule-based model?

5. Is the tool and/or software you describe publicly available?

6. There are a number of minor typographical and grammatical errors throughout the manuscript which require

inspection. There is also a tendency to switch tones - from formal to less formal in places.