Reviewer's report

Title: Languages for different health information readers: Multitrait-multimethod content analysis of Cochrane systematic reviews textual summary formats

Version: 0 Date: 20 Dec 2018

Reviewer: Mike Conway

Reviewer's report:

BMC Medical Research Methodology

Languages for different health information readers: Multitrait-multimethod content analysis of Cochrane systematic reviews textual summary formats

20th Dec 2018

SUMMARY
This article describes work on analysing the readability (and other linguistic characteristics) of summary formats for Cochrane systematic reviews. The text summary types consisted of *press releases (in English), *scientific abstracts (in English), and *plain language summaries in various languages, including English, French, and Croatian. Linguistic tools used for analysis included LIWC, Stanford sentiment analysis tools, and IBM Watson Tone Analyzer, in addition to SMOG, a widely used readability metric.

The researchers found that all summary types were relatively difficult to read, especially given that the recommended reading level for health communication is 6th grade, but that plain language summaries were the easiest to read document type.

This is an interesting and generally well-written paper on an important topic (i.e. appropriate health communication), but it does have several problems (alluded to below).

COMMENTS

p2. abstract. "In all text formats, 'sadness' was the most dominantly perceived emotional tone" - Is this not related to the subject matter (illness) rather than tone per se?

p2. General point: There is a comprehensive literature on hedging in scientific writing (i.e. similar to your "Tentative" category). It is probably a good idea to cite some of this work.

p2. Results. "All text formats had low readability, ranging from a median of 15.6..." need to specify what metric you're using here
p2. Results. "Clout" probably needs a couple of words of explanation

p2. "6th Grade level" - terminology regarding grade levels differs from country to country (e.g. in some countries kids will be older, in others younger), so it's probably important to specify which country you are referring to. I'm assuming you mean the US as the widely used readability measures typically use US grade levels.

p3. "…and that patients and general population" typo/grammar

p3. "Scientific abstract is the basic form of summarising…” typo/grammar

p4. Why did you use the SMOG metric only? There are a bunch of different readability metrics out there, each with its own biases. One way of addressing this involves using several metrics and taking the mean score

p4. The way you have chosen to describe the three levels of data analysis for the IBM tool as emotional, psychological, and writing is a bit confusing, as the first two of these could be reasonably described as psychological. Also, the "psychological" dimension seems to be alluding to Big 5 traits. I'd suggest using Big 5 terminology here.

p4. "LWIC uses pools…” LWIC typo; not sure what "pools" means in this context

p5. This isn't really a criticism, but I'm just interested in why so many of the summaries are written in Croatian (compared to say, French or German — languages that have a greater speaker population)?

p6. "Scientific abstracts had significantly lower analytical tone that the other format" I think it would be helpful to spell out what "lower" and "higher" mean here.

p6. What was your motivation for analysing Big 5 traits? Also, I understood that these kind of tools (e.g. Watson Tone Analyzer) are intended to extract psychological variables from a single writer, and I suspect that most of these summaries are written by a committee. What impact do you think this has?

p7. It's not clear to me how the summary variables you use (analytical thinking, authenticity, etc.) map to Liwc categories as shown at this URL: http://lit.eecs.umich.edu/~geoliwc/LIWC_Dictionary.htm

p8. "The low readability of different textual formats for presenting finding of evidence synthesis in systematic reviews has important implications for the efforts in effective textual health evidence translation to different audiences." This is a bit difficult to understand

p8. "may adversely effect accuracy" what do you mean by "accuracy" here?

p8. "truthful information“ I suggest not using the word "truthful" here.
p8. "the will like it and have confidence..." suggest reformulating

p9. "We used three different deep language processing tools" LIWC cannot be reasonably
described as a deep language processing tool.

p9. "and the proposition to introduce more human, emotional element to enhance the
understanding of basic health information" wording here is a bit strange.

p10. "The data sets used and/or analysed during the current study are available from the
corresponding author on reasonable request" What does "reasonable request" mean in this
context?

Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

Yes

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Yes

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

Yes

Are you able to assess any statistics in the manuscript or would you recommend an
additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further
assessment in your comments to the editors.

I recommend additional statistical review

Quality of written English
Please indicate the quality of language in the manuscript:

Needs some language corrections before being published
**Declaration of competing interests**

Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

No competing interests.

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal