Author's response to reviews

Title: Are decision trees a feasible knowledge representation to guide extraction of critical information from randomized controlled trial reports?

Authors:

    Grace Y Chung (gigigraceyc@gmail.com)
    Enrico Coiera (e.coiera@unsw.edu.au)

Version: 2 Date: 30 July 2008

Author's response to reviews: see over
Dear Editor,

We would like to thank the reviewers for their useful comments. We have modified our paper in order to address the recommendations proposed by the three reviewers. Attached in this document is an explanation of the modification point-by-point corresponding to each reviewer’s feedback.

Kind regards,

Grace Chung, Ph.D.
Comments for Reviewers

Reviewer: Jian-Yun Nie

1. There may be another related work, which analyzes whether it is possible to identify PICO (or PECODR) elements from abstracts:


This has been added in page 7.

Reviewer: Nancy Wilczynski

Major Compulsory Revisions:

1. Some aspects of the paper may be difficult for the reader to understand because of the terms/jargon used. The author should define the following terms:
   - Decision trees (define this early on), semantic structure, text mining, language processing, patient utilities, chance nodes, and true decision trees.

   We do define and give examples of decision trees in detail from page 8 onwards, and prior to any discussion of the trees in methods or results.

   On page 5, we have now inserted “semantic structure or meaning representation.”

2. Additionally, the authors should indicate the definition they used for classifying studies as randomized controlled trials. For example, were quasi-random methods of allocation acceptable?

   We did not include quasi-random RCTs and now make this clear in methods, p11.

3. The 2nd sentence in the second paragraph under the section titled “Information Overload in EBM” contains an error. ACP Journal Club and Evidence-Based Medicine do not “aggregate and distill RCT outcomes into systematic reviews.” These 2 secondary
journals include reviews that have been conducted by others given that they are published and pass pre-defined methods criteria.

The paragraph has been changed accordingly.

4. In the Methods section of the paper, the years searched in PubMed should be stated.

   On page 10, we have stated: “These abstracts were sourced from 197 different journals dating from 1998 and 2006.”

5. Were the abstracts (the corpus) derived from CONSORT endorsing journals? Since the authors state that “In our work, we attempt to exploit the strict design principles and stringent reporting guidelines for RCTs…” one would think the CONSORT endorsing journals should be sampled as all journals indexed in Medline may not adhere to CONSORT reporting guidelines.

   No, we did not restrict our study to Consort journals as our goal was to investigate whether automatic extraction of decision trees is possible for all papers describing RCTs, even the ones that do not conform to the CONSORT guidelines.

   We agree with the proposition that it is likely that the CONSORT statement will improve the quality of reporting and in principle should make it easier to extract elements of decision trees. However at this point, many journals have not mandated CONSORT guidelines and many older reports of RCTs do not. Yet even without the CONSORT statement, RCTs are still supposed to use strict design principles and to follow the criterion for RCTs.

6. How did the authors determine which key decision tree elements to look for?

   We focus on the key structural elements of the tree in this paper, without which a tree cannot be constructed. As this is the first paper we are aware of in exploring this topic, we are sure that much more work will follow in exploring additional elements that assist building comprehensive trees, and we do address this in the limitations section in the first paragraph.

7. The author appear to have confused “allocation concealment” with “blinding” as the examples they provide for allocation concealment (on page 12) “…(single or double blinded, open label)…” are examples of blinding not allocation concealment.

   This has been changed from allocation concealment to blinding.

8. How and when were the categories for intervention information, population information, and outcome information determined? For example, were they derived a priori or based on the corpus analysis?
There were determined a priori. As this is an initial scoping study and as discussed above in answer to question 6, we are sure that other categorizations or elements can be focused on in future work.

9. Was there a reliability check for I, P, and O assignments?

A team of three annotators carried out the assignments, after training and agreeing on consensus assignments for test documents (p14). We do not report kappa agreements for this stage of the analysis.

10. The meaning of the 1st paragraph on the 14th page is unclear.

We are unclear which paragraph is problematic here.

11. How many items were in the disputed set? (page 14 – results)

This has been updated accordingly.

12. How did the authors determine that unique section headings were chosen at the discretion of the authors? (page 16)

While we know that they were unique, and we assume it was the authors’ decisions to choose the heading, we indeed are not truly sure this was the case. As it is a minor point we delete this phrase.

13. Has the data presented on page 20 under “Study of Decision…” taken from the abstracts or the full-text of the articles?

Abstracts only were examined as indicated in the paragraph: “Of the 21 abstracts examined, 15 abstracts (71%) indicate the primary outcome measures as well as the values ...”

Minor Essential Revisions:

1. First paragraph of the introduction, 3rd sentence, “While there remains…” should add “good quality” just prior to “RCTs are designed…”

Done

2. First paragraph of the introduction, final sentence, “As such, RCTs have a crucial place…” should change the final 2 words from “…clinical trials.” to “…providing evidence of treatment effectiveness.”

Done
3. Provide a reference for the 2nd statement (“In total, there are over 230,000 RCT citation entries in PubMed”) under the section titled “Information Overload in EBM”.

We should have been more precise and said “There are over 200k citation entries in PubMed with the Mesh Heading ‘Clinical Trials.’” This is now rectified.

4. The third sentence in the first paragraph of the same section should be changed to provide the actual year rather than stating “…over the last year.”

*Done*

5. On page 6, 1st paragraph, update the sentence “By 2000, the Cochrane collaboration had produced…” by providing 2008 data.

*We did not carry out this analysis but merely cite it, and are not in a position to carry out a further analysis.*

6. Page 6, define the acronym “GEM”

*Done.*

7. Page 7, move the definition of PICO up to its first occurrence.

*Done.*

8. It is not clear what is meant by the final sentence on page 7, paragraph 2, “Some of this work…”

   This has been rephrased: In some cases, researchers pinpoint factual information within a document by identifying textual passages that follow scientific arguments such as Purpose, Interpretation and Findings.

9. The 4 specific questions to be answered in this study should be moved from the “Methods” to the “Introduction”.

*Done*

10. There is no reference to table 1 in the text of the article.

   *This comment inserted in page 14: The number of abstracts in each subcategory of 455 Group B abstracts is summarized in Table 1.*

11. The references should be reported in a consistent format. Currently there are many inconsistencies (e.g., abbreviation of the journal name is used at times while at other times the full journal title is used; volume numbers appear only some of the time).
The inconsistencies have been addressed.

12. Titles of tables should appear at the top of the table.

Titles now at the top.

13. Should indicate in Tables 4 and 8 that data were derived from the R1 group.

Caption for Table 4 changed to: Table 4: Examples of the patterns that occur in the section headings of structured RCT abstracts of Group A.

Caption for Table 8 changed to: Table 8: Overall population information for RCT abstracts in group R1.

Reviewer Carol Friedman

Major Compulsory Revisions:

1. The paper is misleading due to the title and some of the statements in the paper. First, the study reported on explored the use of decision trees for representing limited information from RCTs - although such a knowledge structure is an important step prior to extraction, it still may not be possible to develop an automated method that maps the information in the literature into this structure and therefore the title is misleading. A title such as "Decision Trees: A Knowledge Structure for Representing Critical Information in Randomized Controlled Trials" would be more appropriate.

It is not our intention to ‘mislead’ but equally this paper reports our efforts towards a goal of automated extraction of trees – whether it is ultimately possible or not should not preclude our ambition to address this research problem nor identify it in the title. We modify the title slightly but do wish to retain the statement of our research goal. Our new title is ‘Are decision trees a feasible knowledge representation to guide extraction of critical information in randomized controlled trial reports’

2. The conclusions definitely are not supported by the Methods nor are the Results. The study found that the decision tree elements were identifiable by manual review, signifying that decision trees could be used to represent limited information in RCTs; it does not follow or even suggest that automated extraction would be possible.

It is our goal in this paper to identify if such elements are manually identifiable by human experts. We do not think that have made any strong statement beyond that i.e. nowhere do we state that it is proven we can now extract such elements automatically. We have added the word ‘manually’ in the first sentence of discussion to avoid this misinterpretation, if that is the source of the reading, as well as in the conclusion section of the abstract.
3. Another problem is that related work in knowledge representation is not discussed; for example, it is not clear why a decision-tree is a better representation than a frame-based, logic-based, or graph-based representational system. The authors claim that the experimental design makes RCTs amenable to representation as decision trees but they do not explain why other methods aren't or in what ways decision trees are better than other representational structures; there should be a discussion comparing methods.

We do motivate our choice of decision trees in the section ‘decision trees – macro structures representing RCTs’ at some length.

The frame-based, logic-based or graph structure discussed by the reviewer are lower level data structures which can be used in the implementation of decision trees and so we do not see them as equivalent representations. Decision trees are a ‘macro structure’ in linguistic terms, and have implied semantics e.g. about populations, interventions etc that are not present in the semantically neutral knowledge representations suggested here, and it is this semantic narrative that comes with trees that is appealing. A decision tree may thus be implemented at the underlying data structure level as a frame with semantic constraints and with logical reasoning. Using decision trees do not preclude the use of other methods for representations at the lower levels.

Minor Essential Revisions:

1. pg 10 - The author's searched PubMed to obtain RCT abstracts. From this they could compute precision to see how many were genuine but they could not assess recall using this method and therefore classification of RCT's may be a better subsection title than identifiability.

The reviewer is correct that we do not compute recall, but as the term identifiably is not as precise a notion, would prefer to keep it.

2. pg 11- The groups (R,N, R1, etc.) are difficult to follow - some mnemonics and a table would be useful.

We have improved readability by adding in the label associated with group names when it appears in a text section, and in captions. If a summary table is helpful we would be happy to reaper one, but our changes may be sufficient.

3. pg 15 - The authors report on the number of structured abstracts found – since the subheadings varied considerably it is possible that they missed some structured abstracts - a limitation should be included that the method for identifying structured vs. unstructured was not evaluated.

This point is correct and we added a comment in the Limitations section.
4. pg 23 - The authors state that from the standpoint of automated processing – it will be necessary to use the full text for complete decision trees in many cases. This is conjecture - and may be much more difficult. The authors should back up this claim by manually reviewing some complete articles.

*This paper is confined to the study of abstracts. A sound and systematic manual review of full texts is beyond the current scope of this work, and represents a significant new undertaking.*

*We emphasize in our present study simply that it is at least necessary to use full text even though it is more difficult. The paragraph has been modified: “From the standpoint of automated processing, it will at least be necessary to use the full text for complete decision trees in many cases, particularly if all decision trees with respect to all endpoints or assessments are desired. However, it is beyond our scope to assess the difficulty of extracting complete decision trees from the full text...”*