Questions and Answers

1. To what extent do you think the ideal solution is to standardize randomized control trial (RCT) reporting? I know this is happening to some extent (but likely in varied and inconsistent ways themselves).
   Answer: Standardizing reporting of research will help but one needs to recognize there are many layers and many facets for standardizing research reporting. Standardizing reporting of the protocol and methods, standardizing reporting of the results (data and analyses), standardizing reporting of the interpretation of findings, standardizing the set of concepts to report for human interpretation, standardizing the electronic data exchange for machine interpretation.

2. Can small group systematic reviewers do/practice this automation? If so, how?

3. Why aren't journals unifying the reporting of certain items? A standardized format to require registering certain information that can be informing and (easily) categorized later?
   Answer: Great question, presenters will be touching on this today.

4. Is it realistic eventually to only use CENTRAL for systematic reviews on RCTs (and no longer search MEDLINE, Embase and abandon other search techniques such as screening of reference lists)?
   Answer: The best thing to do (which is what this has in common with many of these tools) is to do your own mini evaluation with your own data and see whether the studies you expect to find are there. CENTRAL is very comprehensive but doesn't claim to have everything.

5. James, how about those non-RCT studies? How much progress has been made so far and where can I learn about it?
   Answer: The Epistemonikos platform identifies systematic reviews, and that last pipeline I talked about based on Microsoft Academic is deliberately designed to be study design (and discipline) neutral. Being able to use the RCT classifier gives RCT identification a huge advantage, it must be said, but there's no reason these tools can't work with non RCTs.

6. Currently what is the cost range of these software products on market?
Answer: They range from free to use, to paid-for software-as-a-service - e.g. you can use Trialstreamer and Epistemonikos for free, use the results (i.e., the studies identified) in the Cochrane pipeline in CENTAL - probably for free, and EPPI-Reviewer for free if you're doing Cochrane / Campbell reviews. We have to charge something to keep it online and pay for staff who provide support.

7. How can journals help improve the accuracy automation of systematic reviews in future publications?
Answer: I'm sure Brian could give a very full answer here, but the best thing would be fully structured publication in addition to the version for human consumption. We want concepts labelled consistently, and data included in machine-readable form.

8. I was thinking about the idea of the original authors of the studies serving as the "master" so that other databases draw from original information as much as possible.
Answer: Yes, if we develop standards that are EASY for original authors to implement to report their data, then everyone else involved (reviewers, editors, systematic reviewers, reporters, policymakers, decision-makers) will find it much easier to USE the data and to INTERACT with the original authors to clarify the findings, and this would improve feedback and expectations to result in better research.

9. How much of this performance is due to the NCTID (clinical trials identifier) in the paper?
Answer: We only use the title and abstract, not the full text, we don't use the NCTID Number in the training/evaluation, and when we test on examples that are missing links the performance only degrades slightly, so we are confident that it has nothing to do with the NCT Number in the text.

10. This is great! Can this be replicated for WHO International Clinical Trials Registry Platform (ICTRP), considering that clinicaltrials.gov might restrict geographically? I understand that current Food and Drug Administration Amendments Act (FDAAA) regulations restrict mandatory reporting to approved products, can this produce biases in these automated processes?
Answer: Around half of the trials in CT.gov are not in the US. I think it can and should be developed for ICTRP too. We started with CT.gov because it is familiar to us. We are mostly focused on safety in new drugs and regulatory science where the results are often more available. There is still lots to do.

11. What about the problem of too many different systematic reviews on a topic (number of original RCTs can be few), many of which are poorly done? It seems there is an assumption that more SRs is always better rather than only high quality, non-duplicative systematic reviews.
Answer: Yes! That’s one of the problems we are trying to address directly.

12. What about automation for journals in different languages?
Answer: A standard for computable expression (machine interpretation) can represent the human language used as a codable concept and can encode the data in non-language-dependent forms. This would also facilitate translation to other human languages.

13. Is it possible to do direct comparison of our systems?
Answer: That would be a great idea. I can ask Shifeng Liu who is looking after it to see if he can manage it.

14. Byron, when we look at Risk of Bias (RoB) 2.0 and Quadas 2 etc. the tools label signaling questions instead of just the risk of bias. Would that data improve the performance as it reflects the signals where information to inform the RoB assessment can be based on?
15. Is it possible to manually edit the results of RoB and produce an updated map (figure) through RobotReviewer?
Answer: Not yet - we have released the updatable study system on our github but haven’t made it work live yet.

16. When you complete the RoB exercise in robot reviewer, what is the final product? Do we get the RoB table and the figures?
Answer: Yes, that is correct, the website gives the table and figures, and this is downloadable as a Word doc, or in JSON format.

17. Is the latest RoB Cochrane tool going to be implemented in RobotReviewer in the future?
Answer: Currently we lack the data to do it; we have relied on the existence of tens of thousands of articles labelled with version 1 of the RoB tool; once version 2 has been used more extensively manually some automation could be possible.

18. On the trip database, as I understood, RobotReviewer was used on abstracts and not full text to label the risk of bias, is that still true? As that would be a different approach, I’m unsure how valid the results would be.
Answer: We do a simplified RoB assessment on abstracts, which predicts the probability of being at low RoB for three domains. This works surprisingly well from the abstract alone, and for this simpler task is actually more reliable than using the full text for the more precise breakdown task. We have published the evaluation of this here https://academic.oup.com/jamia/article/27/12/1903/5907063. We mainly use this score for ranking articles.

19. Standardized reporting seems to go back even further to agreement on standardized outcomes of research. How can that be accomplished?
Answer: The COMET Initiative is one effort to define Core Outcome Measures in Effectiveness Trials -- https://www.comet-initiative.org/

20. There was mention of a tool or risk calculator for likelihood of an SR becoming out-of-date. Where might I find that?
Answer: I think it was a reference to this paper: Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. Annals of Internal Medicine 2007; 147: 224-233

21. Does the systematic review community also need to consider its own formats? For example, the work supporting a published systematic review is not often available in tables using standard ontologies or in a machine-readable format.
Answer: Yes, I would say so. The evidence-based medicine (EBM) on Fast Healthcare Interoperability Resource (FHIR) standards would work for clinical trials and systematic reviews. I focused on clinical trials, as they are the precursors to systematic reviews. There is also a related project Clinical Guidelines on FHIR to create standards for computable Clinical Guidelines.

22. Iain – I’m still not sure that how trip presents RoB is what you would want as here: https://www.tripdatabase.com/evidencemaps/search?criteria=migraine as if there is a linear scale of RoB. All the system does is predict if something is low or not within probability values,
which does not tell something about the actual risk of bias. Thus, a high score for low risk of bias is a valid one but I would be very careful about anything that would not be low risk to present as a scale.

Answer: My favourite way is the presentation on our own site at trialstreamer for the reason you mention. We make our predictions/outputs available to anyone but don’t necessarily endorse how they use them. Also, to be fair, that page on Trip is experimental/proof of concept and labeled as such.

**Resources:**
EBMonFHIR (https://confluence.hl7.org/display/CDS/EBMonFHIR) and COVID-19 Knowledge Accelerator (https://confluence.hl7.org/pages/viewpage.action?pageId=97468919) will be covered later in the session and are developing standards that can make this true or nearly true. Hope is coming.

Our group has a web tool to link CT.gov trials to pubmed articles: http://arrowsmith.psych.uic.edu/cgi-bin/arrowsmith_uic/TrialPubLinking/trial_pub_link_start.cgi - It is live now.