# Auto\_labeling App testing feedback

The first round of testing of the labeling app was conducted by five researchers. During the exercise, some issues that were raised to seek clarifications while some were raised at the end. Key of these issues are listed below:

1. Testers were unclear as to what to do in situations where there is a possibility of selecting more than one sentence in a multi-arm trial.
2. They were also not certain what to do even in a single-arm trial but the information on intervention was spread across multiple sentence.
3. Testers sought clarification on choosing the most appropriate sentence given the fact that interventions will be mentioned in multiple sentences in an abstract. (e.g. as shown in this feedback: “*I've also had some cases where you need multiple sentences together to determine the comparison (e.g. one sentence will say people were randomized to intervention or control, and the intervention is described in a different sentence*”).
4. It was unclear to most what constitutes an appropriate phrase/clause.
5. A need for an appropriate example to consult was raised.
6. Some selected sentences to establish RCT alongside another to show comparison of interventions.
7. In some case, because the intervention has been earlier mentioned, it is abbreviated in the sentence for comparison).
8. The flexibility on the number of articles to evaluate per round was also raise...with popular opinion favoring users deciding how many they have time for per round.

Above are some of the key issues that came up during the testing exercise.