Some Discussion on the

Elmer (3/19)

I think that it is critically important to know what questions we are planning to ask before compiling the database of data to be used to answer these questions. The questions that we are planning to ask will determine what and how we collect the data. At this point, I can think of the following questions, I am sure that others will think of many more:

1. Categorizing criteria -- i.e. developing a set of categories into which guideline criteria fall; in other words a classification scheme for guideline criteria.

2. Developing a "test set" of criteria to be used for many reasons (e.g. developing a criteria language, or perhaps testing existing guideline formalisms against the test set). In order to accomplish this, we have to have some reason to think that the test set is representatives. The fact that it comes from a random sample of guidelines helps, but we might also want some other data, e.g. covers the gamet of categories. We might use the categories from the AMIA 2000 paper. Note that the NGC will not include clinical trials --> this is still an important guideline category.

3. Helping develop/refine a guideline classification scheme. With this database we may be able to make assertions re: different kinds of guidelines based on the kinds of criteria they contain. E.g.

4. What kinds of data items the criteria operate on, with implications toward computerization of guidelines (i.e. are the data items likely to be found in EMRs, what are their data types, etc.).

5. Are there natural divisions defined by criteria that may determine "units of information" that should be coded independently as subguidelines?

These are the questions that I can think of off the top of my head.
As I said, there may be many more.

My second point is that we have to think of how we want to make assertions re: these questions? Are we going to make statistical arguments? Is this feasible with the number of people that we have available to code guidelines?

A third point is that of inter-observer agreement. Can we provide evidence that he data we collect is "correct"? I.e. when we say that A is a criteria that breaks down into a,b,c,d,e but not f,g,h --> is there a better way of validating this than inter-observer agreement?

What about implicit criteria? For example, a breast cancer guideline may assume that patients are post-menopausal females without explicitly stating this in the text (or the title). How do we make sure that these are coded into the database? Should they be coded into the database? If electronic guidelines are to interact with each other, they must express implicit as well as explicit criteria (are these all appropriately put in the "eligibility criteria" bin?).

Aziz' answer (3/19)

I agree with what you say. I think our study plan is reasonable right now. To summarize:

1. We will only encode criteria right now and identify data elements needed by a criterion
2. We will not attempt to characterize data items in this phase of the study.
3. We should all try and do one guideline by mid-week, compare our abstractions/databases and have a discussion about it.
4. If all looks good, we can move on the second guideline.

In other phases, we can attempt to answer many questions including those you have identified below. I understand these subsequent studies and their results will be limited what we do now. Issues about inter-observer agreement are important. In the subsequent studies, we can attempt to minimize those by using models to guide further encoding. For example, we can use the HL7 to characterize data items.

Do you have specific suggestions on how we can change the study? If not, lets do one gl each and see what we learn from that.

Mor's answer (3/19)
I think that we should regard this stage of the criterion database as a pilot study. I don't think that it is a lot of work (Probably not more than a single full work day). I think that point no. 3 in Aziz' email is very important. It will help us to develop the database scheme after some experience with looking at guideline criteria.

Elmer is right in all of his points: (1) for a real study we must consider the goals of the study. However, I consider this as a preliminary pilot study that will educate us on how to design the real study. Then, we can have more guidelines that cover the spectrum of guidelines, and do inter-observer reliability studies on constrained value sets.