Annotation Guidelines:
Clinical Trial Cohort Selection

Amber Stubbs, July 17, 2017

1. Overview

The goal of this annotation task is to match patients to inclusion and exclusion criteria. Each patient has a set of longitudinal records, which will be evaluated against the selection criteria. Annotators will examine the records to determine whether each patient described meets each of the criteria. These annotations will then be used to match patients to the appropriate clinical trials.

2. Patient records

The data set for this task contains records for 296 patients, with 2-5 records per patient. These records have been de-identified, and all the dates in the records have been changed. For each set of patient records, assume that the most recent record is "now", regardless of the year, day, or month actually recorded. Evaluate the previous records according to that interpretation. For example, if the most recent record for a patient has the date May 6, 2153, and the previous record has the date December 20, 2152, and you are evaluating a criterion with the time value of "at least 3 months", you will be able to use these two records to determine the status of that criterion.

In many cases, these patient records have gaps between them, sometimes of years. For the sake of this project, assume that the patient's status does not change between records, until indicated in a subsequent record.

3. Clinical trial selection criteria

We have generated a set of selection criteria, based on criteria for real clinical trials but modified to fit our patient population and records. For each criterion, the annotators will examine the files for evidence that the criteria is met or not met, and annotated the relevant phrase(s) accordingly.

Tag	Criteria
DRUG-ABUSE	Drug abuse, current or past
ALCOHOL-ABUS	Current alcohol use over weekly recommended limits

ENGLISH	Patient must speak English
MAKES-DECISIONS	Patient must make their own medical decisions
ABDOMINAL	History of intra abdominal surgery, small or large intestine resection or small bowel obstruction
MAJOR-DIABETES	Major diabetes-related complication. For the purposes of this annotation, we define "major complication" (as opposed to "minor complication") as any of the following that are a result of (or strongly correlated with) uncontrolled diabetes: • Amputation • Kidney damage • Skin conditions • Retinopathy • nephropathy • neuropathy
ADVANCED-CAD	Advanced cardiovascular disease For the purposes of this annotation, we define "advanced" as having **two or more** of the following: • Taking two or more medications to treat CAD • History of myocardial infarction • Currently experiencing angina • Ischemia, past or present
MI-6MOS	Myocardial infarction in the past 6 months
KETO-1YR	Diagnosis of ketoacidosis in the past year
DIETSUPP-2MOS	Taken a dietary supplement (excluding Vitamin D) in the past 2 months
ASP-FOR-MI	Use of aspirin to prevent myocardial infarction
HBA1C	Any HbA1c value between 6.5 and 9.5%
CREATININE	Serum creatinine > upper limit of normal

4. Annotation

For each criterion, the documents should be assessed to determine whether the conditions are met or not met. If they are, the annotators should annotate the part (or parts) of the text that provides evidence for their assertion. For example, if the criterion is "history of intra-abdominal surgery", then any mention of that procedure being performed for that patient will be relevant and can be annotated. However, a "myocardial infarction in the past 6 months" must have **occurred** within the 6 months from the date on the most recent record, just not be mentioned in a recent record.

If it seems that one record in the patient's history confirms a criterion, but another refutes it, the

evidence can be annotated and marked with the "possible" attribute. However, only use this if the evidence is truly unclear.

The only tag that **requires** multiple annotations is ADVANCED-CAD. For a patient to meet that criteria, they must have any two of the listed symptoms/diseases. Both pieces of evidence must be marked.

For other tags, if there is more than one piece of evidence for a criteria being met, those pieces **can** be annotated, but it is not required.

If there is no evidence of a criterion in the document at all, you may create a "non-consuming tag" (use the "NC elements" drop-down menu to select the tag that you want), and indicate whether the criterion is met or not. In most cases, no mention of a criterion will indicate that it is not met (e.g., if the person's HBA1C is never measured, we cannot say that they met the criteria.). However, in some cases we may assume the opposite: if it is not mentioned whether a person can speak English or make their own medical decisions, this is probably because they **can**, and we can annotate that those criteria are met.

5. Annotation software

Annotation will be done using MAE (Multi-purpose Annotation Environment). The newest test version is available at your account on the UAlbany server, packaged with the test annotation files and these instructions. On most systems you can run the .jar file by double-clicking, loading the DTD, then loading the file you want to annotate. Best practice is to run the software from the command line (see the included user guide).

6. A note on HIPAA/PHI

While the files we are sharing with you have been de-identified and changed to protect the identities of the patients and doctors, we ask that you not share these files or send them to anyone. If you know someone who wants access to the files, please direct them to Amber Stubbs (stubbs@simmons.edu). If you find information in the files that seems to be original PHI, please contact Amber immediately.