Simplifying Clinical Trial Eligibility Criteria
Smart Patients, Inc.
August 30, 2013

Motivation

As patients and their caregivers become more involved in treatment decisions, they are increasingly learning about clinical trials as treatment options. To make it easier for them to search for relevant trials, we propose a simplified and standardized format for eligibility criteria that computer systems can parse and analyze. We expect that software developers would then create better tools for patients and caregivers -- as well as for clinicians and researchers -- to search for and understand clinical trials. This, in turn, could result in more interested trial participants who are also more informed about which trials may be appropriate for them.

Analysis of Current Criteria Definitions

We have analyzed the inclusion and exclusion criteria for all the cancer-related trials from clinicaltrials.gov that are recruiting for interventional studies (approximately 8,500 trials). Among these trials, there are over 1,000 ways to indicate the patient must not be pregnant during a trial, and another 1,000 ways to indicate that the patient must use adequate birth control.

In addition to such variation, a criterion is sometimes specified as an affirmative expression (e.g., pregnant or breastfeeding); in other trials it is specified negatively (e.g., may not be pregnant or breastfeeding). Both forms are used interchangeably in the inclusion and exclusion sections such that there is no consistent difference between the two sections. Clarity of language varies considerably. We also found many transcription errors ranging from simple typos to incorrect units for biometric measurement ranges.

Following are a few of the tens of thousands of unique strings representing inclusion or exclusion criteria in our sample:

- signed informed consent
- ovarian cancer patients
- histologically confirmed uterine sarcoma
- serum creatinine < 1.25 x uln
- good hepatic check-up (total serum bilirubin < 1.5 x uln; ast or alt < 2.5 x uln)
- ecog performance status of 0-1

While the variation may be overwhelming, we found that most criteria are variants on a small number of patterns. Specifically, our analysis suggests the vast majority of criteria generally fall into one of four categories:

- A clinical attribute (pregnancy, ovarian cancer, solid tumor).
- A demographic requirement (age, sex).
- An administrative or behavioral requirement (consent, ability to travel or speak English).
- A biometric measurement (weight, creatinine clearance, tumor size).

A Domain-Specific Language for Eligibility Criteria

A domain-specific language (DSL) is a computer language that's targeted to a particular kind of problem, rather than a general-purpose language that's aimed at any kind of software problem (Martin Fowler, Domain-Specific Languages, 2011). DSLs are often used in software engineering to make it easy for non-programmers to define some aspect of a system. They are especially useful when it is important that the specification be intelligible to both human-readers and computer systems.

A simple DSL would help to meet the main requirements of the eligibility criteria section:

1. The eligibility criteria section of a trial definition should be intelligible by non-medical professionals as well as by medical researchers and clinicians.
2. The language and terms used in the eligibility criteria should be consistent across trials.
3. It should be possible for an application to parse the eligibility criteria to automatically determine whether a patient meets the criteria.

While a more rigorous design process than what we've done thus far will be needed to specify a comprehensive DSL, we have begun to fit the most common criteria into a consistent form that is both readable and computable. Rather than start with a specification, we'll share some examples introducing additional conventions as the complexity of the requirements increases.

Of the thousands of pregnancy and contraception examples, all the meaningful distinctions can be captured in a simple sentence that begins with the phrase *must not* or *must*.

- **Must not** be pregnant.
- **Must** use adequate contraception.

A software application could turn these simple criteria into a series of questions that the patient answers with a *yes* or a *no*.

- Are you pregnant (yes or no)?
- Do you use contraception (yes or no)?

Many trials have stricter requirements on the method of contraception. Additional requirements that do not materially affect the meaning of the principle requirement could be added using a comment or annotation.

- Must use adequate contraception; ***double-barrier contraception, condom with spermicidal jelly, foam suppository or film diaphragm with spermicide, or male condom and diaphragm with spermicide.***

This same format can be modified slightly to specify other clinical or administrative requirements.

- **Must be willing** to travel to the trial location.
- **Must give** signed informed consent.
- **Must not be** HIV positive.
- **Must have evidence** of lung cancer.

For this last example, we can also standardize and parameterize clinical terms so that they can be recognized by software.

- Must have **evidence of** lung cancer.
- Must have **diagnosis of** metastatic disease.
- Must **not have evidence of** brain metastases.

We can also add modifiers to clinical terms that an application could recognize.

- Must have **histologically or cytologically confirmed** metastatic breast cancer.
- Must have **radiographically assessable** disease.

A large proportion of all eligibility criteria relate to biometric measurements and acceptable ranges. We can standardize the expression syntax used for measurements.

- Must have **ECOG between 3 and 5**.
- Must have **Karnofsky performance status > 75%**.

We can recognize synonyms for multiple comparison operators and ranges to allow multiple standard representations.

- Must have Karnofsky > 75%.
- Must have Karnofsky status ≥ 75%.
- Must have Karnofsky **>=** 75%.
- Must have Karnofsky **greater than** 75%.
- Must have **75% >** Karnofsky **> 50%**.
- Must have Karnofsky **between 50% and 75%**.

Where the criteria can be expressed using multiple, interchangeable conventions, the variations can be combined.

- Must have karnofsky performance status > 75% **or** ECOG 0-1.

Many similar quantifiable requirements can be specified using these conventions and a software application could recognize variables, retrieve the value from a health record or lab report and evaluate the expression.

- Must have bilirubin less than or equal to 1.5 x uln.
- Must have life expectancy of at least 18 months.
- Must have platelet count > 90 x 10^9/l.
- Must have absolute neutrophil count ≥ 1500/µl.
- Must be age 18 or older.

Some requirements might be qualified or only apply to certain categories of patient.

- Must use adequate contraception **(females only)**.
- Must have BMI less than 30 **(except Asians)**.
- Must have BMI less than 24 **(Asians only)**.

To specify more complex criteria, simple requirements can be clarified with additional modifiers.

- Must use adequate contraception **defined as any one of**
  - double-barrier contraception condom with spermicidal jelly foam suppository.
  - film diaphragm with spermicide.
  - male condom and diaphragm with spermicide.
  - 100% abstinence.
- Must have adequate bone marrow function **defined as all of**
  - hemoglobin ≥ 8 gm/dl.
  - white blood cells (WBC) ≥ 3000/µl.
  - absolute neutrophil count ≥ 1500/µl.

Low Hanging Fruit

While the above is shared as a non-comprehensive work in progress, the immediate practical point is that significant value can be achieved with a modest amount of standardization along these lines. Our analysis of cancer trials suggests that approximately 20% of all criteria are represented by the following topics:

- Pregnancy, breastfeeding, contraception
- Life expectancy
- Brain metastases
- Kidney function, creatinine, creatinine clearance
- Marrow function, absolute neutrophil count, hemoglobin, platelets, leukocytes
- Liver function, alanine aminotransferase, aspartate aminotransferase, alkaline

phosphatase, total bilirubin
- ECOG or Karnofsky performance status
- Current or prior treatment with specific drug or drug class
- History of heart disease, heart failure, myocardial infarction, angina, unstable angina

As we move toward standardizing eligibility criteria, we can initially focus on the above short list which will improve the ability to search for and understand many trials.

Following are examples of how eligibility criteria for the above topics can be standardized:

- Must not be pregnant or breastfeeding.
- Must use contraception.
- Must not have brain metastases.
- Must have creatinine ≤ 1.25 x uln.
- Must have creatinine clearance > 50 ml/min.
- Must have normal renal function.
- Must have normal marrow function.
- Must have life expectancy ≥ 3 months.
- Must have absolute neutrophil count ≥ 1500/µl.
- Must have hemoglobin ≥ 9 gm/dl.
- Must have platelet count ≥ 75,000/µl.
- Must have leukocytes ≥ 3000/µl.
- Must have alanine aminotransferase < 1.5 x uln.
- Must have aspartate aminotransferase < 2 x uln.
- Must have alkaline phosphatase <= 2.5x uln.
- Must have total bilirubin ≤ 1.5 x uln.
- Must have must have normal hepatic function.
- Must have ECOG performance status 0-3.
- Must have Karnofsky performance status > 75%.
- Must not have prior treatment docetaxel.
- Must not have current treatment with mtor inhibitor;
- Must not have prior treatment with MTOR inhibitor within the last 6 months.
- Must not have history of heart disease including:
  - heart failure
  - myocardial infarction
  - unstable angina.

Please contact Kevin Lawrence at kevin@smartpatients.com with comments and questions.