Analysis of Eligibility Criteria Complexity in Clinical Trials

Jessica Ross, MD, MS[1] Samson Tu, MS[2] Simona Carini, MA[3] Ida Sim, MD, PhD[3]

[1] Dept of Psychiatry, Veteran's Administration Medical Center, San Francisco, CA; [2] Stanford Center for Biomedical Informatics Research, Stanford University, Stanford, CA [3] Division of General Internal Medicine, University of California San Francisco, CA;

Abstract

Formal, computer-interpretable representations of eligibility criteria would allow computers to better support key clinical research and care use cases such as eligibility determination. To inform the development of such formal representations for eligibility criteria, we conducted this study to characterize and quantify the complexity present in 1000 eligibility criteria randomly selected from studies in ClinicalTrials.gov. We classified the criteria by their complexity, semantic patterns, clinical content, and data sources. Our analyses revealed significant semantic and clinical content variability. We found that 93% of criteria were comprehensible, with 85% of these criteria having significant semantic complexity, including 40% relying on temporal data. We also identified several domains of clinical content. Using the findings of the study as requirements for computer-interpretable representations of eligibility, we discuss the challenges for creating such representations for use in clinical research and practice.

Introduction and Background

Clinical trials are one of the most valuable sources of evidence on the efficacy of treatments in humans. Evidence based medicine seeks to apply findings from clinical trials to better evaluation and treatment in clinical populations across all domains of medicine. Aside from a JAMA study that examined the prevalence of exclusion criteria in excluding Women, children, the elderly, and those with common medical conditions from trials, however, there has never been a formal study of the semantic and syntactic features and of the complexity of eligibility criteria[1]. What is needed is a clear specification of the kinds of subjects studied in a trial, i.e., the clinical phenotype studied, to facilitate data pooling and the application of pooled and study-specific results to individual patients or to populations.

While disease diagnosis (e.g. ICD-9) is often used as a stand-in for clinical phenotype, clinical phenotypes used in clinical studies are much more complex than can be captured in an ICD term. The phenotypes studied are often refined by their severity, associated complications, or response to specific treatments (e.g., labile Type II diabetes requiring insulin treatment). These complex phenotypes are explicitly stated in the eligibility criteria for the trial. In fact, the aggregate set of eligibility criteria for any one trial defines the overall clinical phenotype of the population being studied in the trial. Each trial may have dozens of eligibility criteria, and each criterion may be extremely complex, in terms of both grammar and content.

Several groups have created formal representations of eligibility criteria, which include CDISC's ASPIRE[2], Arden Syntax[3], SAGE[4], and GELLO[5]. In related work, our Trial Bank Project team at UCSF has developed the Eligibility Rule Grammar and Ontology (ERGO) and its related formalism ERGO Annotation to standardize complex clinical phenotype descriptions as "templates" that, when combined with terms from ontologies and standardized vocabularies, can reproducibly describe phenotypes in a generic computable representation[6].

The evaluation and development of eligibility criteria representations would be greatly assisted by an understanding of the types and range of clinical and semantic features commonly seen in eligibility criteria. To our knowledge, however, there has never been a formal study of the semantic and syntactic features and the complexity of eligibility criteria. The goal of this study is to analyze a large number of randomly chosen eligibility criteria from actual trials across all clinical domains, to characterize the range and types of complexities present, and to identify common semantic patterns.

Materials and Methods

On December 30, 2003, we downloaded all studies in ClinicalTrials.gov as a collection individual XML files (n=9117), of which 1000 were randomly selected for this study. We extracted the text of these criteria, without any metadata (e.g., designations as "exclusion" or "inclusion" criteria, or clinical domain of study) because we were analyzing only the semantic and clinical content in the criteria themselves (available in Supplementary Materials).

We imported the criteria into an Excel spreadsheet and analyzed them by hand (JR) for their content and

semantic structure. Initially we designated criteria as "comprehensible and selective" or "incomprehensible or non-selective." We designated criteria as "comprehensible and selective" if their logic was readily apparent and effectively defined potential subjects as eligible or ineligible for the study (i.e., stated an allowed or disallowed phenotype). These criteria may have had minor grammatical or spelling errors that did not compromise the underlying logic, and are referred to in the remainder of this text as Comprehensible and Selective Eligibility Criteria(C&S criteria). We designated criteria as "incomprehensible or non-selective" if they were truncated or otherwise nonsensical, or if they did not define a phenotype (e.g., the statement "History of breast cancer allowed" does not describe a phenotype to be excluded or included in the study and is therefore non-selective). We recorded the reasons that criteria were classified as "incomprehensible" or "non-selective."

C&S criteria were then classified along several axes. First, they were designated as "simple" or "complex." Simple criteria were those consisting of a discrete clinical concept expressed as a single phrase (e.g. lung cancer, anemia, uncontrolled hypertension) or its negation (not pregnant), or a simple quantitative comparison (e.g., WBC > 5000 cells/mm^3). If we were able to rewrite a criterion as a simple criterion by omitting a modifier or modifiers without changing the meaning of the criterion, it was counted as simple (e.g., "No uncorrected hypokalemia" was rewritten to "No hypokalemia", the simple negation of a noun phrase). All other criteria were deemed to be complex criteria.

The second axis of classification for the criteria was by their high-level clinical content, i.e., whether the criterion concerned: 1) a clinical attribute of the study participant (e.g., a symptom, a disease), 2) a treatment or intervention on the participant, or 3) a behavior of the participant. These three categories were defined after preliminary examination of criteria suggested them to be mutually exclusive and exhaustive in covering the high-level clinical content of all eligibility criteria.

After classification of criteria by their complexity and their high-level clinical content, we analyzed the criteria for semantic and clinical patterns and tallied the proportion of their presence. We identified and tallied the broad semantic patterns according to the following heuristics.

1. Boolean criteria include those with AND and OR connectors, or parentheses, commas, "/", or other grammatical proxies for AND and OR. If a "with" could be converted to an AND without any loss of information (e.g., "Cirrhosis with a history of ascites" was rewritten to "Cirrhosis AND a history of ascites"), the criterion was counted as 'Boolean'.
2. Partially specified lists were treated as OR statements including each element of the list and therefore designated as Boolean(e.g."History of cardiac disease (e.g. MI, CHF))" was expanded to "History of cardiac disease OR MI OR CHF.
3. Phrases with connectors "without", "except", "unless", and "other than" were counted as 'exclusion connectors'.
4. Criteria with semantic connectors commonly used in clinical eligibility criteria, including "caused by", "defined by", and "documented by", "diagnosed by", and "confirmed by" were tallied as 'delimiting connectors'.
5. Phrases were counted as having 'temporal connectors' if they contain temporal descriptors or comparators (e.g. "at least 3 times per week"), or references to temporal events during the study (e.g., "at hospital discharge").
6. Criteria that could be rewritten as "if then" statements were counted as 'If then' statements (e.g., "Prior sentinel node biopsy allowed provided nodes are pathologically negative" was rewritten to "If prior sentinel node biopsy then nodes are pathologically negative").
7. C&S criteria were counted as 'able to be reduced to simple criteria' if they could be decomposed into only simple statements through Boolean, if..then, or exclusion decomposition(e.g. No history of CHF and CAD can be broken down into the two simple rules: "No history of CHF" AND "No history of CAD").

Furthermore, we identified patterns of clinical content and tallied the proportion of their presence:

1. Criteria pertaining to demographic data.
2. Criteria that refer to informed consent or a participant's willingness to adhere to protocol requirements (e.g., return for follow up).
3. Criteria pertaining to laboratory and other test results and other quantitative data, including stages,, grades or other standardized clinical scales. Also included in this group were criteria containing references to arithmetic "normal values" (e.g. AST > 4 x Upper Limit of Normal).
4. Temporally related criteria. These included criteria having temporal connectors and comparators, and ranged from criteria that

clearly referred to an attribute, intervention or behavior that was present at the "start of the study", to criteria that refer to "within the past 6 months", to criteria that refer to "prior" or "concurrent" attributes, interventions or behaviors without any further specifications We distinguished three subtypes of temporally related criteria:

a) Well specified temporal criteria -- These criteria listed a clear time period with reference to a discrete date (e.g. study day 1, the day the patient received treatment).

b) Moderately specified temporal criteria -- These criteria listed a clear time period, but with no reference to study dates (e.g., within the last 6 months, within the prior 2 weeks),

c) Loosely specified temporal criteria -- These criteria mainly used "prior" or "current" as temporal specifiers.

5. Criteria that implicitly require clinical judgment, the details of which are not specified, but that the average clinician would understand and find meaningful (e.g., "eligible for statin therapy").
6. Criteria that require information beyond the criterion itself, e.g. from study meta-data or other criteria (e.g. "No evidence of metastases" without specifying the type of primary carcinoma).

In addition to classifying our test criteria as above, we also reviewed the expressiveness of CDISC's ASPIRE, Arden Syntax, SAGE and GELLO to represent the eligibility criteria patterns detected in this study.

Results

Criteria were first divided into those that were comprehensible (C&S criteria) and those that were not (Table 1). The vast majority of criteria (93%) were C&S criteria. Table 1 also describes the proportion of criteria classified as incomprehensible due to three main reasons.

Table 1.Comprehensibility of Randomly Selected Eligibility Criteria	
Total Eligibility Criteria	1000 (100%)
Comprehensible Criteria (C&S criteria)	932 (93.2 %)
Incomprehensible Criteria	68(6.8%)
Reasons Incomprehensible	
1.Incomplete Statement	32/68 (47%)
2.Doesn't exclude subjects	31/68 (46%)
3.Other	5/68 (7%)

Approximately 15% of the C&S criteria were either simple statements or could be rewritten as such (Table 2). Many of these included negation or simple arithmetic comparisons. The remaining 85% of the criteria were designated as "complex" and included a variety of semantic patterns, including 35% with more than one type of the complex semantic patterns listed. However, 8% of these C&S criteria could be rewritten into two or more simple statements through Boolean, if..then, and exclusion decomposition.

Table 2. Semantic Complexity and Variation in CSEC (N = 932)	
Simple Criteria (SC)	
Proportion of CSEC	**139/932 (15%)**
SC with **negation**	**21/139 (15%)**
SC with **arithmetic** comparator	**49/139 (35%)**
Complex Criteria (CC)	
Proportion of CSEC	**793/932 (85%)**
CC with **negation**	**205/793 (26%)**
CC with **arithmetic** comparator	**113/793 (14%)**
CC with **Boolean** connector	**423/793 (53%)**
CC with **exclusion** connector	**25/793(3%)**
CC with **defining** connector	**45/793 (6 %)**
CC with **temporal** connector	**371/793 (47%)**
CC with **if…then** statement	**80/793 (10%)**
CC with 2 or more of the above complex semantic patterns	**280/793 (35%)**
CC that can be reduced to SC through decomposition	**64/793 (8%)**

The high-level clinical content of all C&S criteria(Table 3) referred to patient clinical attributes 71% of the time. Criteria specifying treatments or other interventions accounted for 27% of the criteria, and criteria referring to patient behavior occurred 2% of the time. Criteria containing at least two types of clinical content accounted for 7% of the total.

Table 3. Clinical Content Variation	
1.**C&S criteria** specifying patient clinical Attributes	659/932(71%)
2.**C&S criteria** specifying treatments or interventions participant has received or will receive.	314/932(34%)
3.**C&S criteria** specifying patient behavior.	35/932(4%)
4. **C&S criteria** including at least 2 of above 3 types of content.	72/932(8%)

Table 4 shows detailed findings of clinical patterns in C & S criteria. Demographic criteria, including age and sex, were found in 2.5%, and 3% pertained to informed consent, or to a participant's willingness and ability to participate as specified in the study protocol.

Table 4. Details of Clinical Content Variation in C&S criteria(N = 932)	
C&S criteria pertaining to Demographic Data (N = 24/932, 2.5%)	
C&S criteria pertaining to Patient Abilities and Informed Consent (N = 28/932, 3%)	
C&S criteria with Labs, Studies, and Standardized Diagnostic Criteria (N = 219/932, 23%)	
C&S criteria with Serum or Urine Lab tests	7%
C&S criteria with Radiographic Data	1%
C&S criteria with other lab tests(e.g. Echo, EKG, histology, vital signs)	6%
C&S criteria with accepted clinical diagnostic criteria(e.g. stages and grades)	5%
C&S criteria with labs requiring clinical interpretation(e.g. AST > 4 X ULN)	4%
C&S criteria with Temporally Related Features (N = 371/932, 40%)	
Well specified temporal **C&S criteria**	4%
Moderately specified temporal **C&S criteria**	11%
Loosely specified temporal **C&S criteria**	25%
C&S criteria requiring Clinical Judgement (N = 174/932, 19%)	
C&S criteria dependent on Study Metadata (N = 221/932, 24%)	

Approximately 23% of the criteria specified the results of laboratory tests, other studies, and standardized diagnostic criteria, which ranged from requiring a very specific numerical value for a participant to meet the criteria of the study, to criteria that referred to "normal" values (e.g. "AST < Upper Limit of Normal").

Temporally related features were present in 40% of C&S criteria. In approximately 1/3 of these, the timing of clinical assessments, interventions or behaviors were well to moderately well defined in relationship to the start of the study. In the remainder of these criteria, timing was much less precisely specified (e.g., clinical assessments or interventions "prior" to the study, or happening "concurrently").

Another 19% of criteria depended on the clinical judgment of a clinician (e.g. "No other medical or psychiatric illness that would preclude study compliance") without giving specific indications as to how this assessment should be made. Finally, approximately 24% of the criteria relied on details of the particular study that were not available from analysis of the criterion by itself (e.g. "No evidence of metastatic disease" without an indication of the primary cancer).

Discussion

Our analysis of 1000 eligibility criteria randomly drawn from ClinicalTrials.gov demonstrates significant semantic and clinical variability across criteria. This variability presents challenges to informaticians, researchers, and clinicians. The good news for informaticians designing expression languages is that 23% of C&S criteria are simple criteria, or can be reduced to simple criteria through Boolean , exclusion, and if-then decomposition. On the other hand, 77% of C&S criteria remain complex to evaluate as they contain one or more of the following patterns: 9% of complex criteria involve the use of semantic connectors which are not captured by the current representation languages or coded data, 40% require the definition of temporal constraints (many of which are only loosely specified), 19% require clinical judgment, and 24% require linkage to study metadata. Researchers trying to determine patient eligibility for studies face incomprehensible and ambiguous criteria as well as under-specified criteria requiring clinical judgment or assessments. Furthermore, 7 % of criteria require radiographic, histologic, or EKG data that may not be available in coded format in the EHR. Automation of screening based on these criteria may require natural language processing of narrative documents, with attendant problems of sensitivity and specificity. Clinicians seeking to determine if a study's population is similar to their own patients are equally challenged to understand just what clinical phenotype was studied in a given trial.

In this work, we conceptualize the clinical domain referenced in eligibility criteria to be orthogonal to the criteria's form (e.g., Boolean combination) and data sources (e.g., laboratory test results). We aimed to classify criteria in a manner that would be useful across clinical domains, to drive development of domain-independent eligibility criteria representations that will allow for tools and discoveries made in one domain to be applied to all areas of medicine, including facilitating the automation of matching potential subjects to trials, designing trials to expand on findings from prior

studies, and pooling data across studies for meta-analysis. A disease-specific representation format based on data elements and their values, such as that of the ASPIRE project, cannot satisfy the expressivity requirements revealed in this study. Domain-independent languages with composition capability, such as Arden Syntax, SAGE and GELLO, do not have this limitation.

Additionally, criteria requiring referencing external sources, and those based on ill-defined criteria provide obstacles to the automation of clinical tasks across medical domains. Such criteria have to be disambiguated, de-abstracted, and operationalized in terms of available data. Work on these issues will benefit from an interdisciplinary approach in designing ontologies and data structures that allow access not only to current sources of quantitative laboratory data, but also to high-throughput and imaging data that will one day be used regularly in clinical practice.

One strength of our study is that we analyzed eligibility criteria without restriction to clinical domain. This method likely produced a wider range of variance in criteria than if we had restricted our analysis to specific domains, but is more reflective of the true range of complexity in eligibility criteria. Researchers in a specific field may use criterion patterns that are not used in other fields, but we abstracted common semantic patterns and our categories covered all patterns seen in our sample of 1000 criteria.

A limitation of this study is that all of the criteria we analyzed were taken from ClinicalTrials.gov, which may produce a bias towards criteria from the types of trials found in this database, e.g., more quantitative than qualitative. Other biases may include more well formed eligibility criteria because trials registered in ClinicalTrials.gov may be of higher study design quality than unregistered trials. Finally, eligibility criteria reported to ClinicalTrials.gov may be simplified versions of more detailed eligibility criteria in actual study protocols. Our categorization of complexity and clinical and semantic patterns will need to be tested against criteria extracted directly from study protocols to demonstrate its full applicability.

Given the diversity of these criteria, including incomprehensible and ill defined criteria, it may well be advantageous to consider the formation of clear standards for clinical researchers to follow when writing eligibility criteria. Our work provides an initial framework for considering best practices for expressing the clinical and semantic content of criteria. While best practice criteria may be viewed as initially burdensome, we believe that the benefits of having more clearly written criteria will far outweigh the costs, and will facilitate formalization and optimal use of these criteria as specific phenotypes throughout clinical research and care across all domains of medicine.

Conclusions and Future Directions

There is significant variation in both the content and semantic structure of eligibility criteria in clinical trials. Evaluations of how current formal representations handle the categories of clinical and semantic patterns we have documented will help move the field forward. We believe this characterization of eligibility criteria complexity will also help future research on the development of formalized representations to capture criteria, as well as the automated parsing of criteria into these formalized representations for computational support. Advances in generic formal representations of eligibility criteria will provide the necessary semantic foundation for maximizing the ability of computers to help manage and apply complex clinical phenotypes as defined by eligibility criteria in clinical research.

Supplementary Materials

http://rctbank.ucsf.edu/home/ecanalysis.html.

Acknowledgements

This work was supported in part by grant R01-LM-06780 from the NLM, and a VA Psychiatric Research fellowship that supports Dr. Ross.

References

1.Van Spall H.G. et al..*Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic review*. JAMA. 2007 Mar 21;297(11):1233-40. c. 2008

2. Niland, J. et al.,*ASPIRE: Agreement on Standardized Protocol Inclusion Requirements for Eligibility*. 2007 (available in Supplementary Materials).

3. Hripcsak, G., et al. *The Arden Syntax for Medical Logic Modules*. in *Proc Annu Symp Comput Appl Med Care*. 1990. Washington, DC

4.Tu, S.W., et al., *The SAGE Guideline Model: achievements and overview.* J Am Med Inform Assoc, 2007. **14**(5): p. 589-598.

5.Sordo, M., et al. *Description and Status Update on GELLO: a Proposed Standardized Object-oriented Expression Language for Clinical Decision Support*. in *Medinfo*. 2004

6.Tu,S.W., et al. *A Practical Method for Transforming Free-Text Eligibility Criteria into Computable Criteria*. AMIA Annu Symp Proc. 2009.