

---

## Research and Applications

# Criteria2Query: a natural language interface to clinical databases for cohort definition

Chi Yuan,<sup>1,2</sup> Patrick B. Ryan,<sup>1,3</sup> Casey Ta,<sup>1</sup> Yixuan Guo,<sup>1</sup> Ziran Li,<sup>1</sup> Jill Hardin,<sup>3</sup> Rupa Makadia,<sup>3</sup> Peng Jin,<sup>1</sup> Ning Shang,<sup>1</sup> Tian Kang,<sup>1</sup> and Chunhua Weng<sup>1</sup>

<sup>1</sup>Department of Biomedical Informatics, Columbia University, New York, New York, USA, <sup>2</sup>Department of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing, Jiangsu Province, P.R. China, and <sup>3</sup>Epidemiology Analytics, Janssen Research and Development, Titusville, New Jersey, USA

Corresponding Author: Chunhua Weng, PhD, Department of Biomedical Informatics, Columbia University, 622 West 168th Street, PH-20, Room 407, New York, NY 10032, USA (chunhua@columbia.edu)

Received 7 September 2018; Revised 16 November 2018; Editorial Decision 27 November 2018; Accepted 29 November 2018

### ABSTRACT

**Objective:** Cohort definition is a bottleneck for conducting clinical research and depends on subjective decisions by domain experts. Data-driven cohort definition is appealing but requires substantial knowledge of terminologies and clinical data models. Criteria2Query is a natural language interface that facilitates human-computer collaboration for cohort definition and execution using clinical databases.

**Materials and Methods:** Criteria2Query uses a hybrid information extraction pipeline combining machine learning and rule-based methods to systematically parse eligibility criteria text, transforms it first into a structured criteria representation and next into sharable and executable clinical data queries represented as SQL queries conforming to the OMOP Common Data Model. Users can interactively review, refine, and execute queries in the ATLAS web application. To test effectiveness, we evaluated 125 criteria across different disease domains from ClinicalTrials.gov and 52 user-entered criteria. We evaluated F1 score and accuracy against 2 domain experts and calculated the average computation time for fully automated query formulation. We conducted an anonymous survey evaluating usability.

**Results:** Criteria2Query achieved 0.795 and 0.805 F1 score for entity recognition and relation extraction, respectively. Accuracies for negation detection, logic detection, entity normalization, and attribute normalization were 0.984, 0.864, 0.514 and 0.793, respectively. Fully automatic query formulation took 1.22 seconds/criterion. More than 80% (11+ of 13) of users would use Criteria2Query in their future cohort definition tasks.

**Conclusions:** We contribute a novel natural language interface to clinical databases. It is open source and supports fully automated and interactive modes for autonomous data-driven cohort definition by researchers with minimal human effort. We demonstrate its promising user friendliness and usability.

**Key words:** cohort definition, natural language processing, natural language interfaces to database, common data model

---

## INTRODUCTION

The growing volume of electronic health record (EHR) data<sup>1</sup> promises to enable early estimate of the feasibility and effectiveness of eligibility criteria during the design process for randomized controlled trials and comparative effectiveness research studies.<sup>2,3</sup> Cohort

definition is a critical and yet a rate-limiting step, often subjective with poor feasibility, resulting in expensive protocol amendments or failed recruitment. Data-driven cohort definition is appealing for enabling informed and feasible cohort definitions but requires substantial knowledge of clinical terminologies and clinical data

representations, which are often complex and heterogeneous. Developing explorative data queries for cohort definitions manually is costly, unscalable, and prohibitively challenging for clinical researchers to perform autonomously without technical support.<sup>4</sup> Furthermore, interpretations of research criteria, which can be elusive and ambiguous,<sup>5</sup> and translating them to data queries can be subjective and variable, leading to inconsistent cohort queries across different implementers and compromising the integrity of multisite clinical studies. Inappropriately implemented cohort queries can further result in unrepresentative study populations, study delays, decreased enrollment efficiency, increased costs, underpowered analyses, failed clinical studies, and ultimately compromises the internal validity and generalizability of study results. The advances in natural language processing (NLP) methods and common data models (CDMs) widely adopted in EHR data organization brings opportunities for optimizing eligibility criteria design and implementation,<sup>6</sup> including the development of a natural language query interface to clinical databases for sharable and executable cohort definition,<sup>7</sup> reducing human effort by leveraging automated computer processing.<sup>8</sup>

### Structuring eligibility criteria

Eligibility criteria are largely documented as unstructured free-text, which is not readily amenable to computer processing for automated cohort definition or knowledge reuse and sharing. A number of eligibility criteria representations have been developed.<sup>9</sup> The representations are either expert-driven or data-driven, with the latter being facilitated by text mining methods. Tu et al designed the Eligibility Rule Grammar and Ontology<sup>10</sup> for clinical eligibility criteria and employed it to transform free-text eligibility criteria into computable criteria.<sup>11</sup> Bhattacharya and Cantor<sup>12</sup> proposed a template-based representation of eligibility criteria for standardizing criteria statements. Weng et al<sup>13</sup> leveraged text mining to develop a Unified Medical Language System-like semantic network-based representation for clinical trial eligibility criteria called EliXR. In addition to the above ontology-based and rule-based approaches, machine learning methods were applied for information extraction of eligibility criteria, as performed in EliIE.<sup>7</sup> Meanwhile, methods for standardizing categorical eligibility criteria were developed, such as EliXR-TIME<sup>14</sup> for temporal knowledge representation and Valx<sup>15</sup> for numeric expression extraction and normalization. Despite these efforts, there is barely a solution for directly transforming eligibility criteria text to executable, nonambiguous cohort queries for standards-based clinical databases. The only related work was from Parker,<sup>16</sup> who used medical logical modules for generating executable SQL queries for eligibility criteria automatically, but the method has limited interoperability outside of the targeted institutional clinical data repository, and its adoption is undocumented.

### Observational Medical Outcomes Partnership common data model and cohort definition

Observational databases differ in both purpose and design, reflected in different data organizations and representations, and the terminologies used to describe medications and clinical conditions vary from source to source. The Observational Medical Outcomes Partnership (OMOP) CDM<sup>17</sup> standardizes data representations and accommodates a wide range of distributed data sources, such as administrative claims and EHR data, for robust evidence generation

using Big Data.<sup>18</sup> More than 120 participants from around the world have joined the collaborative with a vision to access a network of more than 1 billion patient records to generate evidence about all aspects of healthcare. Observational Health Data Sciences and Informatics (OHDSI) provides a wide range of tools to enable distributed and interoperable research across institutions, including ATLAS,<sup>19</sup> which allows software professionals to create standards-based cohort definitions. However, in ATLAS, users must manually create concept sets 1 by 1, organize the logic relations, and standardized attribute values, which can be laborious and error-prone. These tasks are usually prohibitively difficult for key stakeholders, including investigators and research coordinators, and their outcomes are subject to user skill and variable interpretations of eligibility criteria.

### Natural language interfaces to databases

Criteria2Query aims to translate free-text eligibility criteria to standards-based executable cohort definition queries. The primary technology is a natural language interface (NLI) to databases (NLDBs),<sup>20</sup> which allows users to access information stored in relational databases by typing requests expressed in natural language (eg, English). Research on NLIs has attracted much attention since the 1980s.<sup>21</sup> The first NLDB, called LUNAR,<sup>22</sup> enabled English queries of a moon rock database. An early attempt of a clinical NLI was developed by Epstein in 1978,<sup>23</sup> allowing physicians to access a melanoma database using English queries. In these early implementations, researchers adopted predefined templates to translate natural language free-text into structured queries. With the advances in artificial intelligence, NLIs have become increasingly robust,<sup>24,25</sup> but their utility and adoption were limited by heterogeneous database structures. In the medical domain, there are few studies on NLIs to patient-level databases. Woodyard and Hamel<sup>26</sup> developed a template-based NLI to a clinical database management system to improve clinical decision making by providing a question-answering system supporting predefined commands. Roberts and Demner-Fushman<sup>27</sup> presented a manual annotation process for natural language EHR questions that could benefit question-answering systems on patient data. To our best knowledge, no system exists yet to support natural language querying of complex cohort eligibility criteria, including in widely adopted CDM-based clinical databases.

### Contributions

Criteria2Query is designed to accomplish 3 goals altogether: (1) to implement a systematic information extraction (IE) pipeline to parse free-text eligibility criteria into a structured and computable representation, (2) to improve the interoperability between eligibility criteria and clinical databases by representing eligibility criteria using the OMOP CDM, and (3) to present a novel NLI to enable clinicians and researchers to define cohorts autonomously.

## MATERIALS AND METHODS

### System architecture and data flow

Criteria2Query uses a modular architecture in which all modules are loosely coupled<sup>28</sup> so that each submodule is independent and substitutable by emerging or more advanced methods, allowing maximal extensibility. It has 3 functional modules (Figure 1): (1) a systematic information extraction pipeline for parsing free-text eligibility criteria, (2) a query formulation pipeline for automatic generation of standardized cohort definitions and (3) output to ATLAS for interactive query review, refinement, and execution.

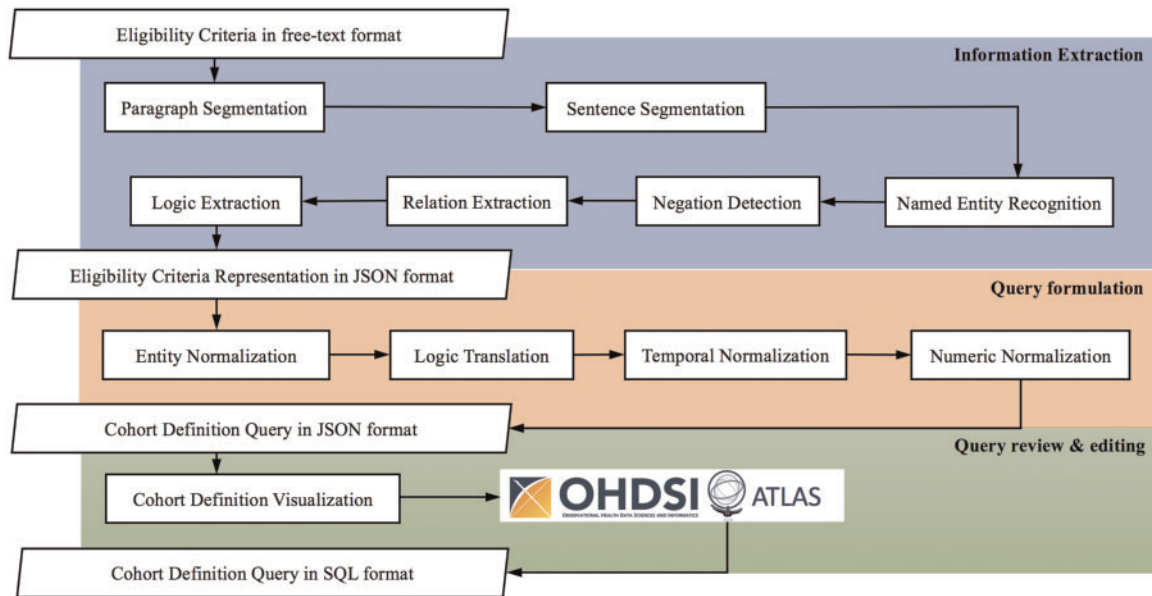


Figure 1. System architecture and data flow of Criteria2Query.

The information extraction pipeline outputs concept-based data representations for all entities accompanied by their negation status, attributes, and logic relations. The query formulation module further processes these representations and outputs OMOP CDM-based cohort queries that can be executed within ATLAS to retrieve patient cohorts satisfying the criteria.

### Systematic information extraction for eligibility criteria

To translate free-text criteria to structured data representations, we developed a systematic IE pipeline with the following steps ordered as paragraph segmentation, sentence segmentation, NER, negation detection, relation extraction and logic extraction. [Supplementary Appendix 1.1](#) shows the information extraction model used in Criteria2Query.

#### Paragraph and sentence segmentation

Generally, eligibility criteria are separated into inclusion and exclusion criteria, each consisting of multiple paragraphs. Paragraphs are separated by line breaks, which are easily recognized. We utilized a sentence splitting method from Stanford CoreNLP<sup>29</sup> to segment sentences, using the default settings. In most cases, sentences and paragraphs are connected by implied “and” logic connections. However, sentences and paragraphs may be connected by different logic, and thus we conducted both paragraph-level and sentence-level segmentation. We implemented a heuristic method to extract patterns from sentences to translate complex logic among sentences (described in Logic detection). For example, the following criterion has complex logic connecting multiple subcriteria: “At least three of the following signs or symptoms of an acute attack of sigmoid diverticulitis must be present: \*Fever (body temperature > 38°C, sublingual), \*Abdominal tenderness, \*Leukocytosis (leukocytes > 10 000/μl) and left shift of the differential blood count (>1% band forms), \*Elevated CRP (> 20 mg/l)” (NCT00097734). Criteria2Query treats the example like a paragraph with the 4 bullet points treated like sentences logically connected by an “AND” operation. We store the pattern information in the paragraph-level pattern element to

indicate that at least 3 of the 4 subcriteria must be satisfied to satisfy the whole paragraph-level criterion.

#### Sentence-level information extraction

**Named entity recognition.** We adapted all annotated criteria from a corpus of 230 Alzheimer’s disease clinical trials provided by a prior publication<sup>7</sup> to fit the latest data representation in OMOP CDM v5.2 by modifying and redefining the categories and attributes of entities (Table 1). We implemented our named entity recognition (NER) methods based on a sequence labeling method, conditional random fields, in CoreNLP<sup>29</sup> with an empirical feature set. After NER, all entities were extracted from free-text criteria with predicted categories assigned automatically ([Supplementary Appendix 1.2](#)).

**Negation detection.** Negation detection is important for determining the negation status for each criterion. We used NegEx<sup>30</sup> with negation trigger files generated previously<sup>13</sup> to assess the negation status of each recognized clinical entity. At the completion of negation detection, every clinical entity is labeled as negated or affirmed. For example, in “No previous myocardial infarction, stroke or diagnosed coronary artery disease” (NCT02834689), the entities “myocardial infarction,” “stroke,” and “diagnosed coronary artery disease” are labeled as negated.

**Relation extraction.** Our pipeline implements binary relation extraction with 2 relationships: has\_temp (temporal) and has\_value (Table 2). Relations between entities are determined by reachability according to enhanced++ English universal dependency parsing results.<sup>31</sup> We implemented a heuristic method by employing Dijkstra’s algorithm<sup>32</sup> to calculate the reachability of each pair of entities. If an entity-attribute pair are connected by a series of modifier relations, then the entity and attribute are recognized as related.

**Logic detection.** The logic operators between entities is crucial for correct semantic representation of eligibility criteria. Hence, we added a logic detection step following the information extraction

**Table 1.** Named entities and attributes recognized by Criteria2Query

	Category	Definition	Examples
<b>Entity</b>	Condition	Conditions are records of a Person suggesting the presence of a disease or medical condition stated as a diagnosis, a sign or a symptom.	<i>Type 2 diabetes mellitus, Alzheimer's disease.</i>
	Drug	Drugs are biochemical substances formulated in such ways that when administered to a person it will exert a certain physiological effect.	<i>Acetaminophen, Furosemide</i>
	Measurement	The standardized examination or testing of a person or person's sample.	<i>Serum creatinine, Serum bilirubin</i>
	Procedure	Procedures are activities or processes on the patient to have a diagnostic or therapeutic purpose.	<i>Chemotherapy, Radiotherapy</i>
	Observation	Observations are clinical facts about a person obtained in the context of examination, questioning or a procedure.	<i>Smoking, drug allergy</i>
<b>Attribute</b>	Value	Numeric attributes include but not limited to age range, lab test result, etc.	<i>30 to 75 years old</i>
	Temporal	Temporal constraints imposed on clinical diagnoses, drugs, etc.	<i>within 12 months</i>

**Table 2.** Relationships in Criteria2Query

Relationship	Entity	Attribute	Example
has_temp	Condition  Measurement  Drug Observation  Procedure	Temporal	<i>"thromboembolic disease" has_temp "within the last 3 months"</i>
has_value	Demographic  Measurement	Value	<i>"Age" has_value "13-15 years old", "platelet count" has_value "&lt; 100 000"</i>

pipeline to resolve the logic operators connecting clinical entities. Our heuristic method uses the conjunct tags in enhanced English universal dependency parsing results<sup>31</sup> to group the entities and decompose the logic relations between entities and groups. A conjunct is the relation between 2 elements connected by a coordinating conjunction, such as *and*, *or*, etc. For example, in the inclusion criteria, "Known to be sero-positive for human immunodeficiency virus (HIV), hepatitis C virus (HCV), or hepatitis B virus (HBV)," all criteria are connected by an "OR" relationship by transitivity of conjunct entities. In a more complicated example, "at risk for GDM (such as having metabolic syndrome, prediabetes, or BMI > 85%; and an A1C < 6.5%)," "metabolic syndrome," "prediabetes," and "BMI > 85%" are all connected by "OR" relationships, and this entire group is connected with "A1C < 6.5%" by an "AND" relationship.

## Query formulation

### OMOP cohort definition

OHDSI's ATLAS tool allows users to manually define cohorts and query OMOP databases. In the cohort definition of OMOP CDM, each criterion has 4 required attributes (Figure 2): inclusion or exclusion, domain (category from Table 1), represented concept set, and temporal requirements. Additionally, relevant attributes (eg, lab results) can be associated with its value. Effective use of the OMOP CDM and ATLAS tools requires substantial experience. Users must review free-text eligibility criteria, create or find suitable concept sets, define criteria 1-by-1, and organize the relations among

criteria. Query formulation in Criteria2Query aims to translate structured criteria (the output from the information extraction pipeline) into OMOP format automatically. Criteria2Query exports criteria definitions as JSON output that can be loaded, visualized, and manipulated in ATLAS.

### Entity normalization

Generating standard concept sets that accurately represent biomedical concepts in free-text is a fundamental but challenging component in query formulation. In free-text criteria expression, entities can semantically represent multiple standard concepts (eg, nonmelanoma skin cancer), but the precise scope of each criterion entity is rarely specified explicitly and requires domain knowledge to define correctly. Reusing pre-existing concept sets defined by experts can increase accuracy and reproducibility. Within the OHDSI community, experts have already created more than 2000 publicly shared concept sets for diseases, drugs and lab tests. Criteria2Query is able to fetch and reuse these concept sets through the OHDSI WebAPI.

We also implemented an automatic concept set generation component to assist users to create new concept sets (Figure 3). As abbreviations are abundantly used in clinical research eligibility criteria, we employed the Unified Medical Language System<sup>33</sup> synonym dictionary to get the full expression of abbreviation terms and map them to vocabularies, for example, extending "AD" to "Alzheimer's Disease." We wrapped a Lucene-based OMOP mapping tool called Usagi<sup>34</sup> as a web service that queries entity terms and their domains to map terms to OMOP standard concepts (Supplementary Appendix 1.3). Using OHDSI application programming interfaces (APIs),<sup>35</sup> we leveraged the rich hierarchical relations among concepts in the OMOP CDM to include all descendants for condition concepts and all drugs sharing the same ingredient for drug concepts.

### Logic translation

The query formulation module takes the concepts and relations produced by the information extraction pipeline, represents them using the concept sets generated in Entity normalization, and formulates query logic using the template introduced in OMOP cohort definition. Given that different CDMs organize and represent logic differently, we developed a logic translation component in Criteria2Query to translate logic within structured criteria to the



**Inclusion / Exclusion:** with at least 1 using all occurrences of:

**Domain:** a measurement of [PHEKB] HBA1c - Concept Set + Add criteria attribute...

**Attributes:** with an abnormal result

**Temporal:** starting between All days Before and All days After event index date  
 \* and ending between All days Before and All days After event index date  
 restrict to the same visit occurrence

Figure 2. An example of one criterion on ATLAS.

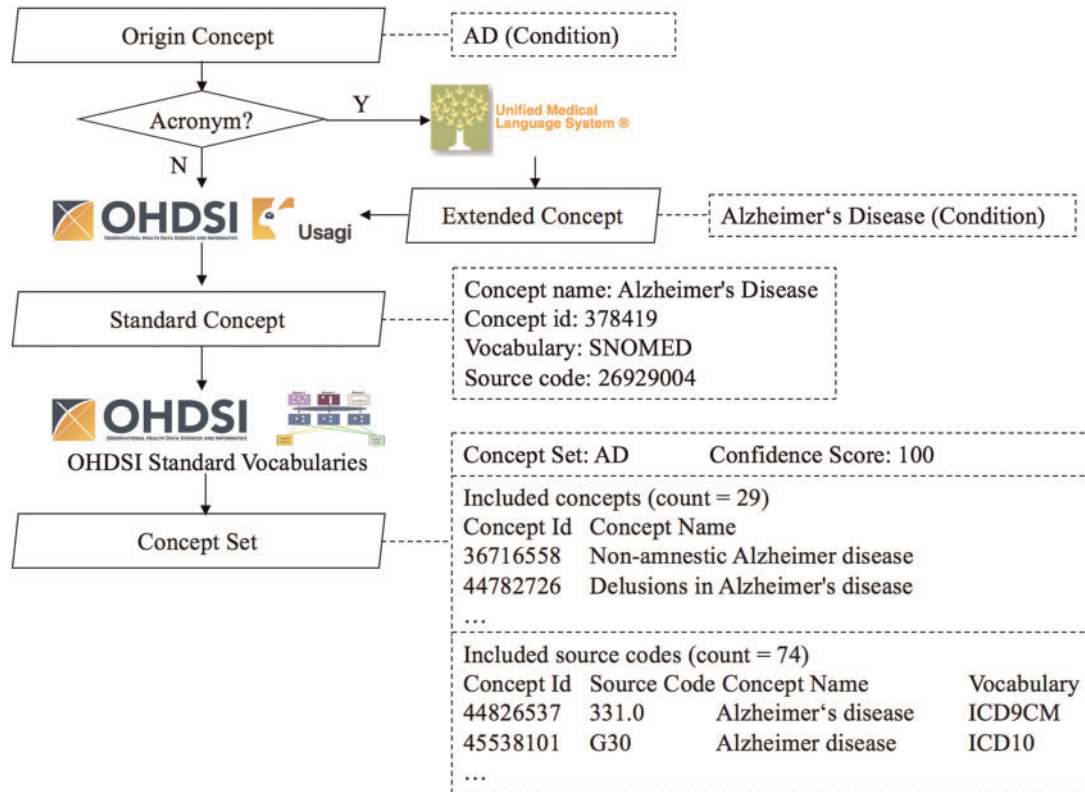


Figure 3. Concept set autogeneration process. AD: Alzheimer's disease; ICD10: International Classification of Diseases–Tenth Revision; ICD9CM: International Classification of Diseases–Ninth Revision–Clinical Modification; N: no; Y: yes.

target data model. In cohort definitions in the OMOP CDM, the logic relations of “AND” and “OR” are represented by the templates “have all of the following criteria” and “have any of the following criteria,” respectively. Exclusion criteria are represented by “with exactly 0 using all occurrences.” Our logic translation component helps users to translate structured logic relations (as produced by the IE pipeline in Section 3.2.2) to logic expressions in the target CDM's cohort definition format. For instance, consider the exclusion criteria, “neurologic disease other than AD” (NCT02167256). This exclusion criterion translates to the definition: “exactly 0 using occurrence” of “neurologic disease” with a subgroup “have all the following criteria” of “Alzheimer's disease.” The logic translation component currently supports the OMOP CDM but is flexible and extensible to support other CDMs.

#### Attribute normalization (temporal and numeric)

Temporal normalization unifies all temporal expressions to the same unit (days). We adapted a library for recognizing and normalizing time expressions, SUTime,<sup>36</sup> to standardized temporal expres-

sions into TIMEX3 format first. We then use regular expressions to transform temporal information in TIMEX3 format into the target CDM format. We also developed a heuristic method for the numeric normalization using regular expressions to fill the results in the target format. Both temporal and numeric attributes are linked to their related criteria based on relation extraction results (Relation extraction).

#### Evaluation methods

##### Evaluation on a random sample of criteria from ClinicalTrials.gov

To test the effectiveness of Criteria2Query on formally written criteria statements, we randomly selected 125 criteria sentences from 10 clinical trials across different disease domains from ClinicalTrials.gov. These 10 evaluation trials were selected outside of the 230 trials previously used to train the system and hence have none overlap with the training data. Evaluation was based on end-to-end results of Criteria2Query. The free-text section of the criteria text block was copied verbatim into the respective inclusion and exclusion criteria text fields in Criteria2Query and automatic processing was per-

formed. Two domain experts provided the gold standards for the indicators. One domain expert reviewed all cohort definitions as visualized in ATLAS, which included the end-to-end evaluation of entity recognition and relation extraction. The other expert reviewed all the concept sets automatically generated by Criteria2Query.

We employed precision, recall, and F1 score to evaluate the performances of the NER and relation extraction components to appraise whether the criteria related entities and relations between entities and attributes were extracted and represented correctly. We also measured accuracies for negation detection, logic detection, entity normalization, and attribute normalization among correctly extracted entities. Entity normalization and attribute normalization were adapted to illustrate the performance of translating free-text expression to OMOP CDM-based structured format. We calculated the 95% confidence intervals for all performance metrics using the adjusted bootstrap percentile interval with 10 000 iterations (R v3.4.4). The computation efficiency was measured by the average time taken for automated query formulation without human intervention.

#### User-centered evaluation method

We conducted an evaluation of Criteria2Query at the 2018 OHDSI annual fall symposium and collected anonymous user feedback and criteria entries from attendees willing to try the demo of Criteria2Query. The study was approved by our IRB as an exempt study. We collected criteria sentences manually entered by OHDSI symposium attendees who tried our software during 2 software demonstration sessions (2.5 hours total). A usage log captured the arbitrary criteria entered by the volunteering testers and their parsing results. Following a brief introduction and demonstration by the CY, the participants had full freedom to test our system without any constraint while entering criteria. We collected all entered criteria from the participants and evaluated accuracy consistently with Evaluation on a Random Sample of Criteria from ClinicalTrials.gov using distinct criteria. Duplicate criteria entries were removed before computation using the aforementioned performance metrics.

We also measured the usability of Criteria2Query. We asked users to take a survey containing 8 questions (Supplementary Appendix 2) after testing our demo and collected the results anonymously on paper. The first 7 questions evaluated each user's prior familiarity with cohort definition and user experience about Criteria2Query. The last question collected free-text comments and suggestions for improvement for Criteria2Query. After we received all paper version surveys, we manually entered these data into SurveyMonkey and reported quantitative analysis of the results automatically generated by SurveyMonkey.

## RESULTS

### User interface and availability of Criteria2Query

Criteria2Query is deployed as a web-based natural language cohort definition system based on the Spring MVC framework. Its online version is available at <http://www.ohdsi.org/web/criteria2query/>. Its source code, test data and evaluation results are available at <https://github.com/OHDSI/Criteria2Query>. The instructions for using Criteria2Query are also on GitHub. Figure 4 shows the user workflow. Users may either enter a ClinicalTrials.gov study ID or free text in the input fields (Figure 5). The "One-Button Start" function (Figure 5) takes users directly to executable queries viewable in ATLAS,

employing auto-generated concept sets and bypassing the intermediate steps. Otherwise, detected entities are highlighted and labeled with their predicted categories. Structured eligibility criteria are downloadable in JSON format. Criteria2Query lists all candidate concept sets for each entity, including the automatically generated concept set and matching concept sets created by domain experts. Interactive entity normalization allows users to select from these concept sets to fine tune the concept mapping results. Finally, the cohort query is ready for review, refinement, and execution for cohort retrieval in ATLAS (Figure 6).

## Evaluation results

### Evaluation results for a random sample of criteria from ClinicalTrials.gov

Criteria2Query was first evaluated on 125 sentences of free-text eligibility criteria, which included 215 entities, 34 relations, 137 negations, and 20 attributes, extracted from 10 randomly selected clinical trials for varying diseases such as Alzheimer's disease, diverticulitis, and lower back pain, from the ClinicalTrials.gov. The full list of NCTID numbers and example cohort queries for the testing criteria can be downloaded from our GitHub repository (<https://github.com/OHDSI/Criteria2Query>) and can be reviewed on the public version of ATLAS ([www.ohdsi.org/web/atlas/](http://www.ohdsi.org/web/atlas/)). The testing criteria cover Demographic, Condition, Drug, Measurement, and Procedure domains of clinical events.

We reported the effectiveness and efficiency of Criteria2Query. In effectiveness evaluation, we designed an evaluation matrix to assess representation performance, and we reported the accuracy of negation detection, logic detection, entity normalization, and attribute normalization (Table 3). The gold standard for all indicators were provided by 2 experts with rich knowledge of medical terminologies and the OMOP CDM. As shown in Table 3, the F1 score for entity recognition and relation extraction is 0.804 and 0.793, respectively. Negation detection, logic detection, entity normalization, and attribute normalization achieved 98.5%, 94.4%, 44.7%, and 80.0%, respectively.

To evaluate the efficiency of Criteria2Query, we assessed the time consumption of the information extraction and query formulation modules. Our experiment environment was a MacBook Pro with Intel Core i7 (3.1 GHz) CPU, 16-GB 2133-MHz LPDDR3 memory, and 512-GB SSD hard disk. On average, each trial only required 15.15 seconds to be translated to an OMOP CDM-compliant structured cohort definition query. Each criterion sentence only needed 1.22 seconds on average. The most time-consuming part of the system is the API call for entity normalization and saving concept sets using the public OHDSI website. A total of 92% of the entire time for generating cohort definition queries was spent on query formulation. Only 8% of the total time was spent on information extraction. A critical NLP task that involves mapping mentions to some standard database or ontology identifiers,<sup>37,38</sup> entity normalization of query formulation turns out to be the most rate-limiting factor in our pipeline because it requires efficient search through vast terminologies for appropriate concept mappings.

### User-centered evaluation results

A pilot user-centered evaluation of the usability of Criteria2Query was conducted during the 2018 OHDSI Symposium. We set up a booth to demo the software and invited conference attendees to try it out. Each user spent 5–10 minutes testing our software. We col-

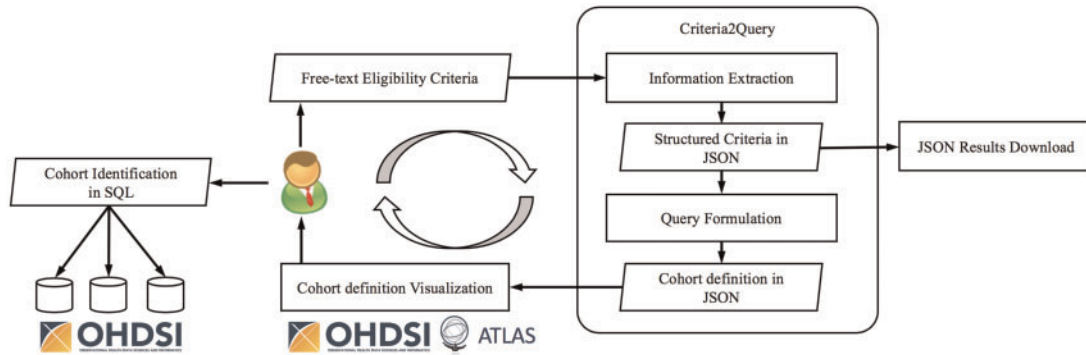


Figure 4. User workflow of Criteria2Query.

Criteria2Query
Support

**✓ Inclusion Criteria**

Tips: Please input criteria line by line

Evidence of the AD pathological process, as confirmed by CSF or amyloid PET scan.  
 Demonstrated abnormal memory function  
 MMSE score greater than or equal to 22 (≥ 22)  
 CDR-Global Score of 0.5 or 1.0  
 Availability of a reliable study partner who accepts to participate in study procedures throughout the 2 years duration of study  
 If receiving symptomatic AD medications, the dosing regimen must have been stable for 3 months prior to baseline and until randomization.

**✗ Exclusion Criteria**

Tips: Please input criteria line by line

Any evidence of a condition other than AD that may affect cognition  
 History of schizophrenia, schizoaffective disorder, major depression, or bipolar disorder  
 History or presence of clinically evident systemic vascular disease that in the opinion of the investigator has the potential to affect cognitive function  
 History or presence of clinically evident cerebrovascular disease

**⚙ Configuration**

Parsing    Reset    One-Button Start

#	Initial Events:	EHR Status
No matching records found		
<b>Inclusion Criteria:</b>		
1	Evidence of the <span style="border: 1px solid red; padding: 2px;">AD</span> <span style="border: 1px solid red; padding: 2px;">CONDITION</span> pathological process , as confirmed by <span style="border: 1px solid red; padding: 2px;">CSF</span> <span style="border: 1px solid red; padding: 2px;">MEASUREMENT</span> or <span style="border: 1px solid red; padding: 2px;">amyloid PET scan</span> <span style="border: 1px solid red; padding: 2px;">PROCEDURE</span> .	YES
2	<span style="border: 1px solid red; padding: 2px;">Demonstrated abnormal memory function</span> <span style="border: 1px solid red; padding: 2px;">CONDITION</span>	YES
3	<span style="border: 1px solid red; padding: 2px;">MMSE score</span> <span style="border: 1px solid red; padding: 2px;">MEASUREMENT</span> greater than or equal to <span style="border: 1px solid red; padding: 2px;">22</span> <span style="border: 1px solid red; padding: 2px;">VALUE</span> ( ≥ 22 )	YES
4	<span style="border: 1px solid red; padding: 2px;">CDR-Global Score</span> <span style="border: 1px solid red; padding: 2px;">MEASUREMENT</span> of <span style="border: 1px solid red; padding: 2px;">0,5 or 1.0</span> <span style="border: 1px solid red; padding: 2px;">VALUE</span>	YES
5	Availability of a reliable study partner who accepts to participate in study procedures throughout the 2 years duration of study	NO
6	If receiving <span style="border: 1px solid red; padding: 2px;">symptomatic AD medications</span> <span style="border: 1px solid red; padding: 2px;">DRUG</span> , the dosing regimen must have been stable for <span style="border: 1px solid red; padding: 2px;">3 months</span> <span style="border: 1px solid red; padding: 2px;">TEMPORAL</span> prior to baseline and until randomization .	YES
<b>Exclusion Criteria:</b>		
1	Any evidence of a condition <span style="border: 1px solid red; padding: 2px;">other than</span> <span style="border: 1px solid red; padding: 2px;">NEGATION CLUE</span> <span style="border: 1px solid red; padding: 2px;">AD</span> <span style="border: 1px solid red; padding: 2px;">CONDITION</span> that may affect cognition	YES
2	History of <span style="border: 1px solid red; padding: 2px;">schizophrenia</span> <span style="border: 1px solid red; padding: 2px;">CONDITION</span> , <span style="border: 1px solid red; padding: 2px;">schizoaffective disorder</span> <span style="border: 1px solid red; padding: 2px;">CONDITION</span> , <span style="border: 1px solid red; padding: 2px;">major depression</span> <span style="border: 1px solid red; padding: 2px;">CONDITION</span> , or <span style="border: 1px solid red; padding: 2px;">bipolar disorder</span> <span style="border: 1px solid red; padding: 2px;">CONDITION</span>	YES
3	History or presence of clinically evident <span style="border: 1px solid red; padding: 2px;">systemic vascular disease</span> <span style="border: 1px solid red; padding: 2px;">CONDITION</span> that in the opinion of the investigator has the potential to affect cognitive function	YES
4	History or presence of clinically evident <span style="border: 1px solid red; padding: 2px;">cerebrovascular disease</span> <span style="border: 1px solid red; padding: 2px;">CONDITION</span>	YES

Next    Download

Figure 5. The user interface of the Criteria2Query system.

**ATLAS**

Initial Event Cohort

People having any of the following:

a visit occurrence of **Any Visit**

with continuous observation of at least **0** days before and **365** days after event index date

Limit initial events to: **all events** per person.

**Initial event inclusion criteria:** From among the initial events, include:

having **all** of the following criteria:

Limit cohort of initial events to: **earliest event** per person.

Remove initial event inclusion criteria

Additional Qualifying Inclusion Criteria

4. [INC]CDR-Global Score of 0.5 or 1.0  
criteria2query v0.8.2.1

5. [INC]Availability of a reliable study partner who accepts to participate in study procedures throughout the 2 years duration of study  
criteria2query v0.8.2.1

6. [INC]If receiving symptomatic AD medications, the dosing regimen must have been stable for 3 months prior to baseline and until randomization.  
criteria2query v0.8.2.1

7. [EXC]Any evidence of a condition other than AD that may affect cognition  
criteria2query v0.8.2.1

8. [EXC]History of schizophrenia, schizoaffective disorder, major depression, or bipolar disorder  
criteria2query v0.8.2.1

9. [EXC]History or presence of clinically evident systemic vascular disease that in the opinion of the investigator has the potential to affect cognitive function  
criteria2query v0.8.2.1

10. [EXC]History or presence of clinically evident cerebrovascular disease  
criteria2query v0.8.2.1

11. [EXC]At risk for suicide in the

Limit qualifying cohort to: **earliest event** per person.

Cohort Exit Criteria

Apache 2.0 open source software provided by OHDSI join the journey

**Figure 6.** Automatically generated cohort query presented by ATLAS to allow query review, refinement, and execution for patient cohort generation using clinical databases.

lected a set of 94 criterion sentences manually entered by 13 OHDSI symposium attendees who tried out our software demo. After removing 42 criteria as duplicates, we retained 52 unique criteria for evaluation. The full list of user-entered criteria and evaluation results can be downloaded from our git repository. As shown in Table 3, the F1 score for entity recognition and relation extraction were 0.775 and 0.833, respectively. Negation detection, logic detection, entity normalization, and attribute normalization achieved 97.9%, 50.0%, 80.8%, and 77.8%, respectively.

All the 13 testers finished our anonymous surveys. The first 3 questions ask about participants' prior familiarity with cohort definition. When asked about their level of experience with self-service tools for cohort definition (eg, ATLAS or i2b2), 15.4% (2 of 13), 61.5% (8 of 13), and 23.1% (3 of 13) of participants responded "no experience," "a little experience," and "very experienced," respectively. Almost half of the participants consider it difficult to perform the task of cohort definition (eg, identifying queryable eligibility concepts, mapping concepts to terminology codes, and translating eligibility logic to database query expressions).

The last 5 questions ask for participants' opinions about Criteria2Query; 100% (13 of 13) of participants either completely or somewhat agreed that the NLI of Criteria2Query was user friendly. When asked if Criteria2Query is difficult to use, 23% (3 of 13) agreed somehow or completely, 15.4% (2 of 13) of participants were neutral, and 61.6% (8 of 13) disagreed somehow or completely. A total of 84.6% (11+ of 13) of participants indicated a willingness to use Criteria2Query in their future cohort definition tasks. Only 4 participants provided free-text feedback, which is generally positive with constructive feedback, for question 8 ("do you have further comments about Criteria2Query?") and the other 9 participants provided no response, as shown in Supplementary Appendix 2.

## DISCUSSION

We present Criteria2Query, a novel NLI for transforming free-text clinical research criteria to OMOP CDM-based cohort queries. It facilitates EHR-based cohort definition using a series of NLP meth-



**Table 3.** The evaluation matrix of criteria representation with 95% confidence intervals

Evaluation Matrix	Criteria crawled from ClinicalTrials.gov (n = 125)				Criteria Entered by Testers (n = 52)				Combined (n = 177)				
	Precision	Recall	F1	F1	Precision	Recall	F1	F1	Precision	Recall	Precision	Recall	F1
Entity recognition	0.902 (156/173) [0.844-0.936]	0.726 (156/215) [0.661-0.777]	0.804 [0.760-0.841]	0.804 [0.760-0.841]	0.899 (62/69) [0.783-0.942]	0.681 (62/91) [0.571-0.758]	0.775 [0.694-0.833]	0.775 [0.694-0.833]	0.901 (218/242) [0.851-0.930]	0.712 (218/306) [0.657-0.758]	0.901 (218/242) [0.851-0.930]	0.712 (218/306) [0.657-0.758]	0.795 [0.758-0.828]
Relation extraction	0.958 (23/24) [0.792-1.000]	0.676 (23/34) [0.471-0.794]	0.793 [0.576-0.867]	0.793 [0.576-0.867]	1.00 (10/10)	0.714 (10/14) [0.357-0.857]	0.833 [0.526-0.923]	0.833 [0.526-0.923]	0.971 (33/34) [0.824-1.000]	0.688 (33/48) [0.521-0.792]	0.971 (33/34) [0.824-1.000]	0.688 (33/48) [0.521-0.792]	0.805 [0.647-0.871]
Accuracy													
Negation detection	0.985 (135/137) [0.944-1.000]	0.942 (135/137) [0.942-0.993]	0.993 [0.942-1.000]	0.993 [0.942-1.000]	0.979 (47/48) [0.896-1.000]	0.979 (47/48) [0.896-1.000]	0.979 (47/48) [0.896-1.000]	0.979 (47/48) [0.896-1.000]	0.984 (182/185) [0.946-0.995]	0.984 (182/185) [0.946-0.995]	0.984 (182/185) [0.946-0.995]	0.984 (182/185) [0.946-0.995]	0.984 (182/185) [0.946-0.995]
Logic detection	0.944 (17/18) [0.722-1.00]	0.722 (17/18) [0.722-1.00]	1.00 [0.722-1.00]	1.00 [0.722-1.00]	0.500 (2/4) [0.000-0.750]	0.500 (2/4) [0.000-0.750]	0.500 (2/4) [0.000-0.750]	0.500 (2/4) [0.000-0.750]	0.864 (19/22) [0.591-0.955]	0.864 (19/22) [0.591-0.955]	0.864 (19/22) [0.591-0.955]	0.864 (19/22) [0.591-0.955]	0.864 (19/22) [0.591-0.955]
Entity normalization	0.447 (51/114) [0.351-0.535]	0.351 (51/114) [0.351-0.535]	0.535 [0.351-0.535]	0.535 [0.351-0.535]	0.808 (21/26) [0.577-0.885]	0.808 (21/26) [0.577-0.885]	0.808 (21/26) [0.577-0.885]	0.808 (21/26) [0.577-0.885]	0.514 (72/140) [0.429-0.586]	0.514 (72/140) [0.429-0.586]	0.514 (72/140) [0.429-0.586]	0.514 (72/140) [0.429-0.586]	0.514 (72/140) [0.429-0.586]
Attribute normalization	0.800 (16/20) [0.500-0.900]	0.500 (16/20) [0.500-0.900]	0.900 [0.500-0.900]	0.900 [0.500-0.900]	0.778 (7/9) [0.222-0.889]	0.778 (7/9) [0.222-0.889]	0.778 (7/9) [0.222-0.889]	0.778 (7/9) [0.222-0.889]	0.793 (23/29) [0.586-0.897]	0.793 (23/29) [0.586-0.897]	0.793 (23/29) [0.586-0.897]	0.793 (23/29) [0.586-0.897]	0.793 (23/29) [0.586-0.897]

Values are Precision, Recall, F1 score (n/n) [95% confidence interval] or Accuracy [95% confidence interval], unless otherwise indicated.

ods and OMOP CDM terminologies. From the usability perspective, compared with manual and form-based methods for SQL query crafting, Criteria2Query has 3 distinctive features.

First, Criteria2Query proposes a systematic information extraction method for structuring eligibility criteria text, promoting knowledge reuse. In previous studies,<sup>7,11</sup> more preprocessing was required since all input could only take the form of sentences. Criteria2Query splits the task at different levels from document to sentences to make the whole parsing process systematic (Systematic Information Extraction for Eligibility Criteria). For informatics researchers, they are able to customize their own logic on different levels. A library of frequently used criteria further enables the development of the next-generation eligibility criteria authoring systems that can be data-driven and knowledge based, optimizing study feasibility and population representativeness, as previously envisioned by Weng.<sup>6</sup>

Second, Criteria2Query extends the open-source OMOP CDM and OHDSI APIs. The OMOP CDM has rich concept representations, which map to many source vocabularies and hence is able to support semantic queries. For example, in our entity normalization module, we can easily map drug ingredients mentioned in free-text criteria to all brand-name drugs and dosages containing that ingredient. For example, “insulin, isophane” could be mapped to a concept set containing “insulin human, isophane 100 UNT/ML Pen Injector,” “insulin, isophane Pen Injector [Humulin N],” and more than 2000 other drug formulations.

Third, Criteria2Query is the first of its kind that takes user input of natural language criteria to generate OMOP CDM-based executable clinical database queries. Researchers can use Criteria2Query in either automated fashion to receive query results using default recommendations, or semi-automated fashion to refine the results at each step. It enables researchers to query EHR data autonomously for cohort definition without requiring them to master medical terminologies or database query languages (eg, SQL). In addition to unstructured eligibility criteria, users may add other standardized fields from ClinicalTrials.gov or other clinical protocols into the Initial Events section for customized results.

Compared with manual cohort definition, Criteria2Query has 4 advantages for minimizing user effort and for standardizing outputs. First, Criteria2Query highlights clinical entities and attributes and labels their EHR presence status automatically, making it easy for users to refine the criteria as needed. Second, pregenerated concept sets could be shared and reused, maximizing knowledge reuse and collaboration. In related systems, such as i2b2, users must specify the clinical concept codes to query their database. Third, Criteria2Query is able to formulate the logic relations according to the target data model, OMOP CDM. Users only need to perform drag-and-drop based on pregenerated cohort definitions in ATLAS instead of starting from scratch. Finally, the units of attributes, including numeric and temporal attributes, are converted to the CDM’s standard formats.

### Open source, flexibility, and extensibility

Criteria2Query is open source, modular, and follows loose coupling design. Compared with other systems, Criteria2Query has the advantages of being flexible and extensible. This design allows users to reuse, interchange, or enhance individual modules as needed without affecting other system components. For example, users who only need the structured criteria representation may employ the information extraction module independently to transform their data.

Users can also extend capabilities by incorporating new modules that conform to a standard interface. For example, phenotyping algorithms using semi-structured data (eg, International Classification of Diseases–Ninth Revision codes) simply need a translation component to leverage the query formulation module to generate phenotype queries automatically. We provide RESTful APIs to facilitate its integration with other systems. Other informaticians can generate SQL queries by Criteria2Query and implement more data visualization of cohorts for executing those queries, which could help them optimize their criteria input.

### Error analysis

The information extraction errors can be attributed to suboptimal entity recognition and relation extraction. The low recall in entity recognition could be due to training our sequence labeling model solely on a corpus of 230 Alzheimer’s disease trials. For example, in the criterion, “Patients to be included in the study must have AMD with choroidal neovascularization,” “AMD with choroidal neovascularization” was not recognized. This situation was exacerbated when the users attempted to search for recognizable concepts repeatedly. For example, during our user evaluation, 31% (4 of 13) of the test users repeatedly tried different variations of the same class of concepts that could not be recognized, each time adding to the count of errors. For example, 1 user tested 4 forms of “patient with a diagnosis of mitral stenosis and is on pitting edema and is on <DRUG>,” with tykysin, tikosyn, and dofetilide being different stenosis drugs.

Relation extraction errors were mainly due to the incomplete sentence structures common in free-text eligibility criteria. As the Enhanced++ English universal dependency parsing<sup>31</sup> was designed for general language dependency parsing, it had difficulty identifying the dependencies in the abbreviated sentence structures common in medical corpora. More training data and rules need to be added to our heuristic methods for future improvements.

Criteria2Query achieved a significantly better accuracy in entity normalization with user entry data than criteria extracted from ClinicalTrials.gov. Most user-entered entities were simple enough to be satisfied by the concept sets automatically generated by Criteria2Query using the hierarchy of vocabularies. The poor performance in entity normalization for ClinicalTrials.gov data was partially due to the complexity and ambiguity of biomedical terms and to the oversimplified automatic mapping of biomedical terms. A medical domain expert reviewed the automatically generated concept sets to identify error causes for the worse results from ClinicalTrials.gov data (Supplementary Appendix 1.4). According to our analysis, the majority of errors in entity normalization were caused by a lack of domain knowledge to understand which concepts should be included in ambiguous hypernyms, such as “nonmelanoma skin cancer” and “anticoagulant.” More than 40% of all entity normalization errors could be remedied by reusing existing concept sets created by domain experts.

Logic detection performed worse in user-entered criteria compared with criteria from ClinicalTrials.gov (50.0% vs 94.4%, respectively), but the difference was not significant. The criteria entered by testers were generally less complex, containing only 4 logic relations among the 52 evaluated criteria. The errors occurred in 2 criteria entered by 1 user with the same nested logic pattern: “patient can not have <Condition A> or <Condition B>, but <Condition A> is ok if no <Condition C>.” The nested logic in “but <Condition A> is ok if no <Condition C>” was not detected by Criteria2Query.

## Limitations and future work

As the initial NLI to clinical databases for cohort definition, Criteria2Query has several limitations. According to these limitations, we outlined its corresponding future improvements.

First, the training set for the conditional random fields-based machine learning model used in our NER module only included Alzheimer's disease trials, while our testing data included trials from any disease domain and contained more diverse criteria expressions. Due to its limited scope, our training data did not contain similar feature patterns as the testing data, causing NER failures. Although our model exhibited some generalizability of capturing common expressions in criteria, performance could further improve with larger and more diverse training data. We are annotating a larger corpus of eligibility criteria with samples from a broad range of clinical trials to train a more robust NER model. In addition, the current version of Criteria2Query supports only offline learning and does not accommodate online learning for unrecognized or misclassified named entities. To address this limitation, we will add an interactive function in our system to allow users to edit the NER results shown in Figure 5. This will enable continuous online learning so that the online corrections to NER results can be incorporated to enhance future NER performance.

Second, more comprehensive entity normalization is required. Eligibility criteria are often vague and ambiguous, using hypernyms or fuzzy terms to represent a set of diseases, for example, "significant medical or psychiatric disorder" (NCT01825512) and "severe or uncontrolled systemic disease" (NCT00807170). These terms refer to broad sets of diseases that may vary with subjective opinion. Given the ambiguity, it is hard to create a concept set accurately representing these terms. In our system, we calculated the string distance to find the most relevant concepts to create its concept set, but this approach may not work well in these scenarios. We will involve domain experts to semimanually create reusable concept sets and promote knowledge reuse by reviewing the concept sets of frequently used criteria in clinical trials. We will also collaborate with the OHDSI community to leverage the Gold standard phenotype library under development (<http://forums.ohdsi.org/t/requirements-development-for-the-ohdsi-gold-standard-phenotype-library/4876>).

Third, Criteria2Query currently does not have an ideal solution to configure initial event cohorts (eg, "patients with Alzheimer's disease"), which is defined by an initial event (eg, "first diagnosis of Alzheimer's disease") and an observational time window (eg, "3 years after the first diagnosis"). However, this information is required in ATLAS for executing the query against a clinical database to generate a cohort. This information determines the anchor event and within what time windows around the anchor event a patient's medical history is considered for eligibility determination, whose configuration challenges vary by different use case scenarios. If eligibility criteria are for prospective clinical trials recruiting patients, there is no information about anchor event and observational windows for Criteria2Query to use. If eligibility criteria are for replicating trials from ClinicalTrials.gov as retrospective studies, an anchor event and an observational window must be manually specified for the study. By default, Criteria2Query uses the observational windows of any visit(s) in the target database as the observational windows and makes any visit as the initial event. One of the potential future improvements is to make Criteria2Query smarter and to enable automatic extraction of the anchor event from structured clinical trial summaries (eg, from the condition or medication information) or automatic inference of the anchor event from user-entered eligibility criteria based on their relatedness to certain diseases.

Finally, in this study, we only conducted evaluation on a small criterion set from ClinicalTrials.gov and a small-scale user evaluation study. Further, conducting the user evaluation during conference demo sessions may not reflect how users would realistically interact with the system for cohort definition. For example, several users entered overly simplistic queries (eg, "dead," "on clopidogrel") possibly due to a combination of unfamiliarity with the system and the transient nature of demo sessions. A large-scale and rigorous evaluation is warranted and underway.

We will evaluate Criteria2Query on a diverse criteria corpus, including complex criteria from proprietary clinical research protocols and large number of user-entered criteria to report the generalizability of the results to a broader set of clinical research criteria. We also realize that jargon such as initial events, index criterion, and attributes may be foreign to some users. Adequate training can help familiarize the users with the functionality and workflow of Criteria2Query. Therefore, understanding the users and designing tailored training materials to reduce learning curves is also necessary before large-scale user engagement. To quantify the time saving of our tool, we will record all operations during the evaluation to analyze the real-world performance of Criteria2Query.

## CONCLUSIONS

Criteria2Query systematically translates free-text eligibility criteria to CDM-based cohort queries. It demonstrates early promise for empowering researchers and clinicians to create patient cohorts without mastering clinical CDMs or query languages by providing a NLI. Future longitudinal user evaluation studies at larger scales are warranted to assess its impact on facilitating clinical research.

## FUNDING

This research was supported by National Library of Medicine grant R01LM009886 (to CW), Janssen grant JANSRD CU15-2317 (to CW), and National Center for Advancing Translational Science grants 3OT3TR002027-01S1 (to CW) and UL1TR001873 (PI: Muredach Reilly) for disseminating the research results to Columbia CTSA.

## AUTHOR CONTRIBUTORS

CY, PR, and CW conceived the system design together. CY designed and implemented the system. CW supervised CY in his design and implementation. CW and CT edited the manuscript critically. CT, NS, ZL, JH, RM, and JP contributed to the design and evaluation of the system. YG and TK contributed to individual modules or prior versions of the system. All authors edited and approved the manuscript.

## LICENSE

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive license (or nonexclusive for government employees) on a worldwide basis to the BMJ Group and co-owners or contracting owning societies (where published by the BMJ Group on their behalf), and its Licensees to permit this article (if accepted) to be published in the Journal of the American Medical Informatics Association and any other BMJ Group products and to exploit all subsidiary rights, as set out in our license.

## SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

## ACKNOWLEDGEMENTS

The authors would like to acknowledge the Observational Health Data Sciences and Informatics technical teams for their kind help, particularly Lee Evans. They also thank James Rogers for his help with the user-centered evaluation.

*Conflict of interest statement.* None.

## REFERENCES

- Häyrynen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *Int J Med Inf* 2008; 77: 291–304.
- Penberthy L, Brown R, Puma F, et al. Automated matching software for clinical trials eligibility: measuring efficiency and flexibility. *Contemp Clin Trials* 2010; 31 (3): 207–17.
- Thadani SR, Weng C, Bigger JT, et al. Electronic screening improves efficiency in clinical trial recruitment. *J Am Med Inform Assoc* 2009; 16 (6): 869–73.
- Penberthy LT, Dahman BA, Petkov VI, et al. Effort required in eligibility screening for clinical trials. *J Oncol Pract* 2012; 8 (6): 365–70.
- Musen MA, Rohn JA, Fagan LM, et al. Knowledge engineering for a clinical trial advice system: uncovering errors in protocol specification. *Bull Cancer* 1987; 74: 291–6.
- Weng C. Optimizing clinical research participant selection with informatics. *Trends Pharmacol Sci* 2015; 36 (11): 706–9.
- Kang T, Zhang S, Tang Y, et al. EliE: an open-source information extraction system for clinical trial eligibility criteria. *J Am Med Inform Assoc* 2017; 24 (6): 1062–71.
- Friedman CP. A “fundamental theorem” of biomedical informatics. *J Am Med Inform Assoc JAMIA* 2009; 16 (2): 169–70.
- Weng C, Tu SW, Sim I, et al. Formal representation of eligibility criteria: a literature review. *J Biomed Inform* 2010; 43 (3): 451–67.
- ERGO: a template-based expression language for encoding eligibility criteria. <http://rctbank.ucsf.edu/home/ergo>. Accessed April 7, 2018.
- Tu SW, Peleg M, Carini S, et al. A practical method for transforming free-text eligibility criteria into computable criteria. *J Biomed Inform* 2011; 44 (2): 239–50.
- Bhattacharya S, Cantor MN. Analysis of eligibility criteria representation in industry-standard clinical trial protocols. *J Biomed Inform* 2013; 46 (5): 805–13.
- Weng C, Wu X, Luo Z, et al. EliXR: an approach to eligibility criteria extraction and representation. *J Am Med Inform Assoc* 2011; 18 (Suppl 1): i116–24.
- Boland MR, Tu SW, Carini S, et al. EliXR-TIME: a temporal knowledge representation for clinical research eligibility criteria. *AMIA Summits Transl Sci Proc* 2012; 2012: 71–80.
- Hao T, Liu H, Weng C. Valx: a system for extracting and structuring numeric lab test comparison statements from text. *Methods Inf Med* 2016; 55 (3): 266–75.
- Parker CG. Generating medical logic modules for clinical trial eligibility. PhD thesis, Brigham Young University; 2005.
- OMOP Common Data Model – OHDSI. <https://www.ohdsi.org/data-standardization/the-common-data-model/>. Accessed March 25, 2018.
- Hripsak G, Duke JD, Shah NH, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform* 2015; 216: 574–8.
- ATLAS. <http://www.ohdsi.org/web/atlas/#/home>. Accessed March 25, 2018.
- Androutopoulos I, Ritchie GD, Thanisch P. Natural language interfaces to databases—an introduction. *Nat Lang Eng* 1995; 1 (1): 29–81. doi:10.1017/S135132490000005X.
- Copestake A, Jones KS. Natural language interfaces to databases. *Knowl Eng Rev* 1990; 5 (4): 225–49.
- Woods WA. Progress in natural language understanding: an application to lunar geology. In: Proceedings of the June 4-8, 1973, National Computer Conference and Exposition. New York, NY: ACM; 1973:441–50.
- Epstein MN, Walker DE. Natural language access to a melanoma data base. *Proc Annu Symp Comput Appl Med Care* 1978; 1978: 320–5. doi:10.1109/SCAMC.1978.679936.
- Chandra Y, Mihalcea R. Natural language interfaces to databases. Master of Science thesis, University of North Texas, 2006.
- Pazos RRA, González BJJ, Aguirre LMA, et al. Natural language interfaces to databases: an analysis of the state of the art. In: Castillo O, Melin P, Kacprzyk J, eds. *Recent Advances in Hybrid Intelligence Systems*. New York: Springer; 2013; 463–80.
- Woodyard M, Hamel B. A natural language interface to a clinical data base management system. *Comput Biomed Res* 1981; 14 (1): 41–62.
- Roberts K, Demner-Fushman D. Toward a natural language interface for EHR questions. *AMIA Summits Transl Sci Proc* 2015; 2015: 157–61.
- Loose coupling. Wikipedia. 2018. [https://en.wikipedia.org/w/index.php?title=Loose\\_coupling&oldid=840712645](https://en.wikipedia.org/w/index.php?title=Loose_coupling&oldid=840712645). Accessed May 14, 2018.
- Manning C, Surdeanu M, Bauer J, et al. The Stanford CoreNLP natural language processing toolkit. In: Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations. Baltimore, MD: Association for Computational Linguistics; 2014:55–60. <http://www.aclweb.org/anthology/P14-5010>. Accessed July 18, 2018.
- Chapman WW, Bridewell W, Hanbury P, et al. A simple algorithm for identifying negated findings and diseases in discharge summaries. *J Biomed Inform* 2001; 34 (5): 301–10.
- Schuster S, Manning CD. Enhanced English universal dependencies: an improved representation for natural language understanding tasks. <https://nlp.stanford.edu/pubs/schuster2016enhanced.pdf>. Accessed March 26, 2018.
- Dijkstra’s algorithm. Wikipedia. 2018. [https://en.wikipedia.org/w/index.php?title=Dijkstra%27s\\_algorithm&oldid=832284506](https://en.wikipedia.org/w/index.php?title=Dijkstra%27s_algorithm&oldid=832284506). Accessed March 26, 2018.
- Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res* 2004; 32 (90001): 267D–70.
- Usagi. <https://github.com/OHDSI/Usagi>. Accessed April 7, 2018.
- OHDSI/WebAPI. [http://webapidoc.ohdsi.org/job/WebAPI/WebAPI\\_Mire-dot\\_Documentation/index.html](http://webapidoc.ohdsi.org/job/WebAPI/WebAPI_Mire-dot_Documentation/index.html). Accessed April 7, 2018.
- Chang AX, Manning CD. Sutime: a library for recognizing and normalizing time expressions In: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*. Istanbul, Turkey: ELRA; 2012:3735–740. <https://pdfs.semanticscholar.org/fe80/746646f6ee819205ca9a8476292ea6b83e66.pdf>. Accessed July 5, 2017.
- Dogan RI, Lu Z. An inference method for disease name normalization. Paper presented at: AAAI Fall Symposium: Information Retrieval and Knowledge Discovery in Biomedical Text; November 2–4, 2012; Arlington, VA.
- Zhou L, Plasek JM, Mahoney LM, et al. Mapping partners master drug dictionary to RxNorm using an NLP-based approach. *J Biomed Inform* 2012; 45 (4): 626–33.