

## Visual aggregate analysis of eligibility features of clinical trials



Zhe He<sup>a</sup>, Simona Carini<sup>b</sup>, Ida Sim<sup>b</sup>, Chunhua Weng<sup>a,\*</sup>

<sup>a</sup> Department of Biomedical Informatics, Columbia University, New York, NY 10032, USA

<sup>b</sup> Division of General Internal Medicine, University of California, San Francisco, San Francisco, CA 94143, USA

### ARTICLE INFO

#### Article history:

Received 29 August 2014

Accepted 12 January 2015

Available online 20 January 2015

#### Keywords:

Clinical trial

Patient selection

Selection bias

Knowledge management

### ABSTRACT

**Objective:** To develop a method for profiling the collective populations targeted for recruitment by multiple clinical studies addressing the same medical condition using one eligibility feature each time.

**Methods:** Using a previously published database COMPACT as the backend, we designed a scalable method for visual aggregate analysis of clinical trial eligibility features. This method consists of four modules for eligibility feature frequency analysis, query builder, distribution analysis, and visualization, respectively. This method is capable of analyzing (1) frequently used qualitative and quantitative features for recruiting subjects for a selected medical condition, (2) distribution of study enrollment on consecutive value points or value intervals of each quantitative feature, and (3) distribution of studies on the boundary values, permissible value ranges, and value range widths of each feature. All analysis results were visualized using Google Charts API. Five recruited potential users assessed the usefulness of this method for identifying common patterns in any selected eligibility feature for clinical trial participant selection.

**Results:** We implemented this method as a Web-based analytical system called VITTA (Visual Analysis Tool of Clinical Study Target Populations). We illustrated the functionality of VITTA using two sample queries involving quantitative features BMI and HbA1c for conditions “hypertension” and “Type 2 diabetes”, respectively. The recruited potential users rated the user-perceived usefulness of VITTA with an average score of 86.4/100.

**Conclusions:** We contributed a novel aggregate analysis method to enable the interrogation of common patterns in quantitative eligibility criteria and the collective target populations of multiple related clinical studies. A larger-scale study is warranted to formally assess the usefulness of VITTA among clinical investigators and sponsors in various therapeutic areas.

© 2015 Elsevier Inc. All rights reserved.

## 1. Introduction

Well-designed clinical study protocols are essential for generating high-quality medical evidence [1]. However, studies are often criticized for lacking generalizability, or external validity [2–16]. Because population representativeness is an important aspect of clinical research generalizability, study designers should justify the tradeoffs between internal validity and external validity that arise from their choices of eligibility criteria. Biased or overly restrictive eligibility criteria may (1) exclude patients who may need or benefit from the research [4,5], and (2) lead to an overestimate of the efficacy of an intervention [10]. For example, according to Schmidt et al. [12], almost none of their analyzed studies on secondary prevention of cardiovascular events justified the applied

exclusion criteria, which excluded 21–97% of the female target population. Similarly, Zimmerman et al. [15] reported that approximately 32–47% of patients with major depressive disorder would have been excluded by two most commonly used cutoff values of the Hamilton Rating Scale for Depression in antidepressant efficacy trials, i.e., 18 and 20.

When designing clinical studies, investigators often reuse eligibility criteria from previous protocols of related studies. One of our previously published papers also discovered that many clinical studies, especially those on the same medical condition, use similar or identical eligibility criteria [17]. Therefore, we hypothesize that the generalizability issue might be not only at the level of individual studies, but also at the community level in the entire clinical trial enterprise. Unlike prior work that looks at the generalizability of one study at a time, we are motivated to assess the collective generalizability by uncovering collective design patterns for participant selection among multiple related clinical trials. Unfortunately, at present there is no method or tool for making such

\* Corresponding author at: Department of Biomedical Informatics, Columbia University, 622 W 168th Street, VC-5, New York, NY 10032, USA.

E-mail address: [cw2384@cumc.columbia.edu](mailto:cw2384@cumc.columbia.edu) (C. Weng).

design biases transparent or help investigate such biases. Echoing this need, recently the National Center for Advancing Translational Sciences (NCATS) responded to the Institute of Medicine's review of the Clinical and Translational Science Awards (CTSA) Program in the United States and identified "lack of a knowledge base for all types of interventions at the extremes of age as well as within special populations" as one of the weaknesses of the current translational science enterprise [18]. To help bridge this gap, a computable repository of eligibility features of clinical trials is needed to analyze the characteristics of the target populations on a large scale [19].

The study and result registry ClinicalTrials.gov [20] created by the National Library of Medicine is a valuable public data source. Since September 27, 2007, all United States-based clinical trials of FDA-regulated drugs, biological products, or devices have been mandated to be registered in ClinicalTrials.gov [21]. As of March 18, 2014, 163,285 clinical studies conducted in more than 180 countries were registered in ClinicalTrials.gov. Study summaries are stored in a semi-structured format in the registry, i.e., study descriptors such as title, phase, and location are organized in structured fields. The eligibility criteria are usually organized as paragraphs of free-text or as bullet lists.

The ClinicalTrials.gov is a preferred resource to be transformed into a computable repository of reusable knowledge of clinical trial designs. However, there is little published work on building a computable repository from study summaries on ClinicalTrials.gov. Tasneem et al. developed the Aggregate Analysis of ClinicalTrials.gov (AACT) database as a publicly accessible analysis dataset derived from ClinicalTrials.gov [22]. Using AACT data, clinical trials in various domains have been systematically analyzed, e.g., infectious diseases [23], oncology [24], and diabetes [25], to name a few. AACT allows selection and aggregation of trials by study descriptors, such as study status, phase, and intervention type, but not by fine-grained clinical characteristics of the target population. As studies often limit eligibility to permissible ranges of quantitative features as age, BMI, HbA1c, and blood glucose level [26], investigators or policy makers may be interested in analyzing such quantitative features across studies addressing the same medical condition, with questions like "what is the range of BMI values that are permitted across interventional studies on Type 2 diabetes?" However, as most of the eligibility criteria are in unstructured text, it remains difficult to support these analyses in a programmatic, accurate and scalable way. Hence, to date, there is a paucity of analyses on the quantitative eligibility features of target populations of existing studies, and consequently a lack of capacity to optimize the eligibility criteria definition for future clinical studies based on past studies.

We have developed methods for parsing eligibility features from free-text eligibility criteria [17,27–41] and the derived frequent eligibility features across ClinicalTrials.gov study summaries have produced promising results for searching and indexing studies [29], probing disease relatedness [30], and clustering studies with similar eligibility criteria [17]. Enabled by these techniques, we have created a database of discrete clinical trial eligibility features extracted from ClinicalTrials.gov called COMPACT (Commonalities in Target Populations of Clinical Trials) [42], which allows users to flexibly query sets of clinical studies (e.g., Type 2 diabetes studies) on their shared eligibility features (e.g., HbA1c or BMI) and attributes (e.g., allowed value range for HbA1c or BMI). In addition, we have developed a distribution-based method for profiling clinical trial target populations across sets of studies [43]. Meanwhile, as one of the state-of-the-art methods for discovering knowledge from Big Data [44,45], interactive visual query interfaces can be employed to further support flexible profiling of target populations of sets of clinical studies and to investigate the generalizability of these studies. It has been used for tasks similar to profiling target

populations, such as visualizing alternative disease progression paths for a group of patients similar to a query patient [46], and for visual analysis of clinical event patterns through a combination of a graphical query interface, pattern mining and visualization techniques [47]. Therefore, we enhanced our COMPACT database of study summaries with visualization of the distributions of sets of clinical studies along any single quantitative eligibility feature. To the best of our knowledge, this effort represents one of the earliest attempts to perform aggregate analyses of clinical trial eligibility criteria design patterns. Fig. 1 illustrates the design of the methodology framework, which integrates text mining, data warehousing, and data visual analytics for rich information made available by ClinicalTrials.gov. This pipeline can help clinical trial designers more easily understand collective design patterns in clinical trial eligibility criteria across multiple related clinical trial studies. On this basis, our system can increase the transparency of hidden eligibility criteria design biases at the clinical research community level. Our system supports flexible study selection using multiple study descriptors, such as study type, study design, intervention type, phase, condition, gender, and age range. We hypothesized that our method could identify understudied population subgroups whose value ranges for certain quantitative eligibility features were systematically excluded or overly researched according to analyses of eligibility criteria specifications. Our preliminary user evaluation confirmed this hypothesis and the value of our method for improving the transparency of clinical trial participant selection decisions.

The remainder of the paper is organized as follows. Section 2 first describes the visual aggregate analysis system of eligibility features of clinical trials and how a user interacts with it, and then delineates the methods used to develop and evaluate the system. In Section 3, we use Type 2 diabetes and hypertension as example conditions to illustrate the functionalities of the system. We also present the results of a preliminary evaluation with a convenience sample of five potential users of the system. Finally, we discuss the implication and the limitations of this work in Section 4 and draw conclusions in Section 5.

## 2. Methods

Previously, we introduced a novel database called COMPACT, which stores metadata and parsed eligibility criteria of study summaries in ClinicalTrials.gov [42]. It supports retrieval of readily analyzable eligibility features, quantitative or qualitative, from sets of studies. On this basis, we designed an interactive visual analysis system to aggregate target populations of sets of clinical studies. The potential users of this system include clinical investigators, study sponsors and policy makers. Table 1 presents the glossary of terms that are frequently used in this paper.

Our system enables a user to select a medical condition, one of the quantitative eligibility features frequently seen in studies on that medical condition and other additional study descriptors (e.g., study type, study design, intervention type) to perform five analyses: (1) distribution of number of studies over consecutive value points or non-overlapping value intervals within user-specified value range of the selected quantitative eligibility feature (e.g., over each 0.5% of HbA1c); (2) distribution of enrollment over those value points or value intervals; (3) distribution of number of studies over boundary values (e.g., lower bound of HbA1c as 7.0%); (4) distribution of number of studies over permissible value ranges (e.g., BMI between 15 and 25 kg/m<sup>2</sup>); and (5) distribution of number of studies over value range widths (e.g., the value range width for HbA1c between 7% and 10% is 10 – 7 = 3). Fig. 2 illustrates the comparison between information provided by the ClinicalTrials.gov and our system called VITTA (Visual Analysis Tool of Clinical

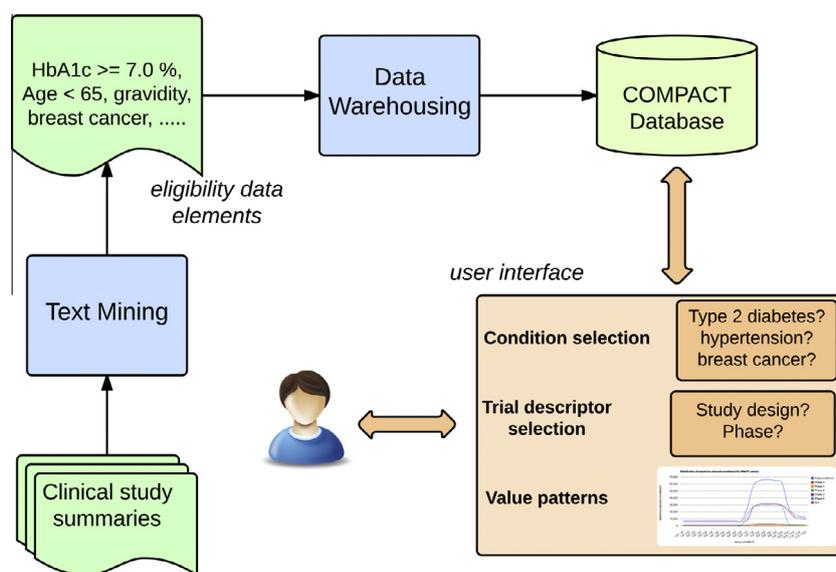


Fig. 1. The methodology framework.

**Table 1**  
The glossary.

Term	Description
Quantitative features	Eligibility criteria with a permissible numeric (continuous) value range requirement for participants, e.g., “BMI < 30 kg/m <sup>2</sup> ”
Qualitative features	Eligibility criteria that accept one of a set of value options, e.g., “with acute ischemic stroke” (yes or no)
Value spectrum	The collective value range of a quantitative feature in selected studies that use the feature for eligibility determination
User-specified value range	The value range of a quantitative feature specified by the user for analyzing the distribution of studies and enrollment. For example, a value range of HbA1c can be “4–12%”
Value intervals	Evenly (based on the user-specified width) or unevenly (based on the actual boundary values) divided smaller bins in the user-specified value range
Permissible value range	The value range specified by an eligibility criterion. For example, the permissible value range for criterion “age between 15 and 40” is [15,40], where “[ ]” means inclusive
Value range width	The difference between the upper bound and lower bound of a permissible value range. For example, the value range width for the criterion “age between 15 and 40” is 25 (=40 – 15)
Boundary value	The threshold value(s) of a quantitative feature in an eligibility criterion, e.g., the boundary values for the criterion “age between 15 and 40” are 15 and 40

cal Study Target Populations). Conventionally users who want to analyze the aggregate target populations of Type 2 diabetes studies in ClinicalTrials.gov may search for studies whose condition is “Type 2 diabetes.” They would then review the summary of each of the studies returned, look for “HbA1c” in the eligibility criteria section, note the permissible value range and finally manually aggregate all the values retrieved. In contrast, our system can automatically group sets of studies and summarize such information from the pre-processed study summaries in the COMPACT database.

The workflow (Fig. 3) for a user is as follows: (1) select a medical condition and specify the number of frequent features to retrieve; (2) select a quantitative eligibility feature among the most frequent quantitative features of the studies on the condition selected in step (1); (3) specify a value range of the selected quantitative feature and optionally one or more study descriptors to further restrict the set of studies; (4) choose a type of analysis on the quantitative feature; and (5) receive the visual query results.

For example, if a user chooses “Type 2 diabetes” as the condition and sets the number of frequent features to retrieve as “10”, the system returns a list of 10 most frequent quantitative eligibility features extracted from Type 2 diabetes studies. From this list, the user may choose “HbA1c.” The user is then asked to specify the value range of interest for HbA1c and additional characteristics for the studies to be included in the analysis, such as study type,

study design, intervention type, or patient age range. The user may formulate a query as “distribution of enrollment in randomized studies in Phase 2/3 on Type 2 diabetes sponsored by industry over value intervals of HbA1c between 1% and 14%.” The width of value intervals can be fixed or varying. If the user specifies a width (e.g., “0.5”), the range 1–14% is evenly divided into value intervals of the same width (e.g., [1%, 1.5%], [1.5%, 2%], ..., [13.5%, 14%]). Alternatively, the user may perform this analysis on varying-width value intervals, which are generated based on the actual boundary values of HbA1c in the selected studies. Hypothetically, this search would include studies NCT00374907 (enrollment: 156) and NCT00097084 (enrollment: 324). The range of HbA1c in NCT00374907 is 6–8% and in NCT00097084 it is 7–11%. These four boundary values (6%, 7%, 8%, 11%) would be used to divide the range 1–14% into five value intervals of varying widths: [1%, 6%], [6%, 7%], [7%, 8%], [8%, 11%], and [11%, 14%]. The system would then analyze the enrollment of the studies that recruit subjects with HbA1c in each value interval and visualize the result. Table 2 shows the distribution of enrollment (anticipated or actual) for this example. Column 4 (maximum enrollment) shows the aggregate enrollment of all the studies in each interval, whereas column 5 (average enrollment) shows the average of study enrollment for all the studies in each interval.

The system also allows the user to visually identify frequently or rarely used boundary values, permissible value ranges, and per-

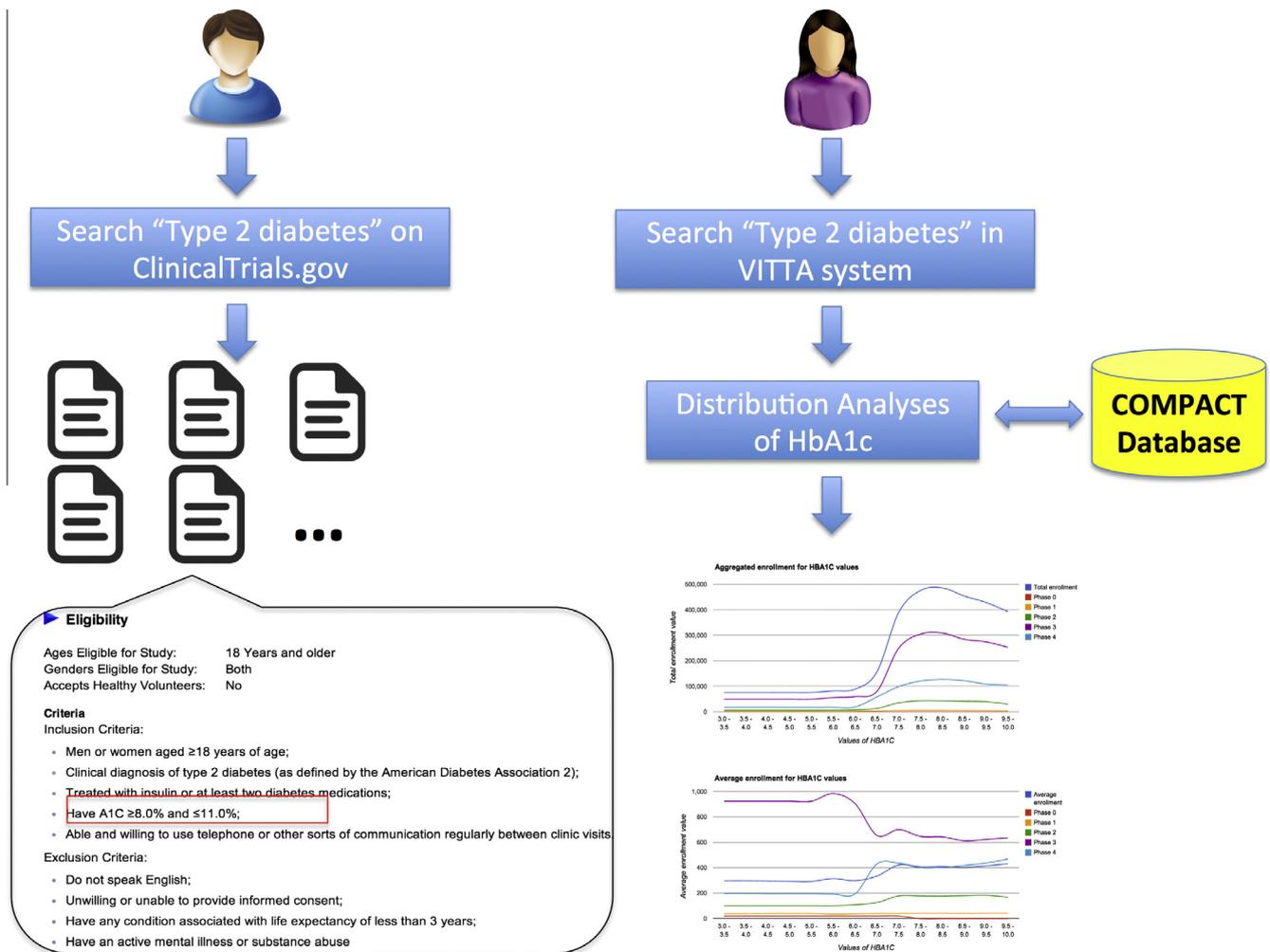


Fig. 2. The comparison between information provided by ClinicalTrials.gov and VITTA.

missible value range widths for the selected quantitative feature. The analytical framework underlying this system comprises four modules: an eligibility feature frequency analysis module, a query builder, a distribution analysis module, and a visualization module, each supporting a step of the user's interaction with the system. In the rest of this section, we first briefly describe how we processed the study summaries in ClinicalTrials.gov for building the COMPACT database. (The design and implementation detail of COMPACT are described in [42].) Then we explain in detail the design and implementation of these four modules.

### 2.1. Parsing study summaries

From the ClinicalTrials.gov website we downloaded the XML records of all the studies and transformed data elements of interest into a computable format using Python programs in four steps: (1) extract study metadata such as study type, intervention type, study phase, gender, and age; (2) identify studies for a certain medical condition; (3) mine frequent eligibility features from free-text eligibility criteria; (4) structure numeric expressions for the frequent quantitative features. Next, we describe these four steps in detail.

#### 2.1.1. Extracting metadata of studies

From each downloaded XML file, we extracted study characteristics and structured the free-text eligibility criteria text using previously published methods [28,48]. We created a table in

COMPACT and saved the extracted metadata of studies: National Clinical Trial (NCT) number (a unique ID assigned by ClinicalTrials.gov), study type (e.g., interventional, observational), intervention type, study design (i.e., allocation and intervention model for interventional studies; time perspective for observational studies), phase, sponsor agency type (e.g., NIH, industry), enrollment status, start date, gender, minimum age, maximum age, and enrollment. From COMPACT, we can flexibly retrieve studies with various characteristics, e.g., *all the observational diabetes studies sponsored by NIH*.

#### 2.1.2. Identifying studies for a certain medical condition

To retrieve condition-specific high-frequency eligibility features, we obtained all the NCT numbers and the list of 4412 medical conditions listed in ClinicalTrials.gov records [49]. We pre-indexed the studies of 1311 conditions, each having more than 50 studies in ClinicalTrials.gov at the point of the study. With the APIs provided by ClinicalTrials.gov, studies using synonyms of the same condition such as "heart attack" and "myocardial infarction" were automatically consolidated [50].

#### 2.1.3. Identifying the most frequent eligibility features

We extended a published method for mining frequent tags [28] to extract frequently used eligibility features from eligibility criteria. A frequent tag is an n-gram completely or partially matching a UMLS concept to a semantic type most relevant to clinical study

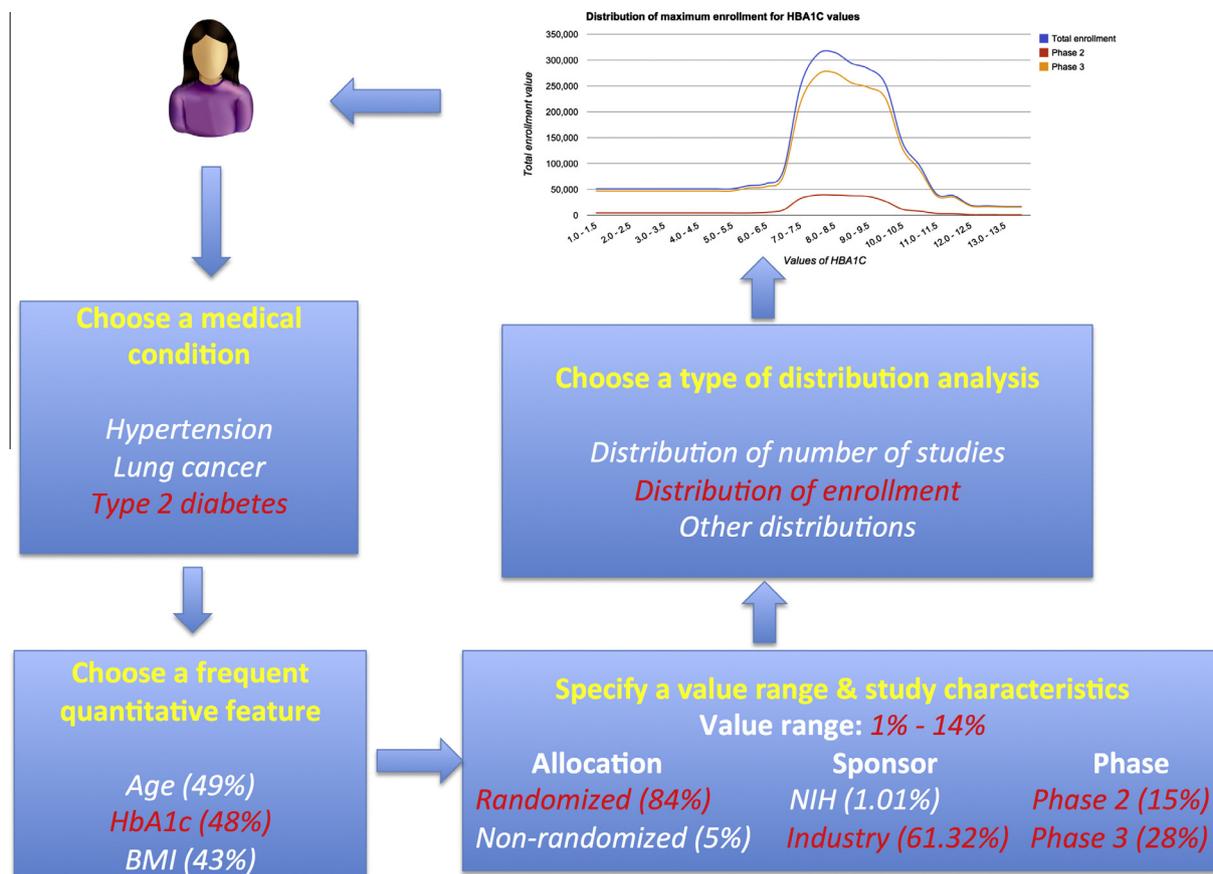


Fig. 3. The workflow for a user to interact with VITTA.

Table 2

Distribution of enrollment of two sample studies on varying-widths value intervals of HbA1c.

Value interval	Included studies	Number of studies	Maximum enrollment	Average enrollment
[1.0%,6.0%]		0	0	0
[6.0%,7.0%]	NCT00374907	1	156	156
[7.0%,8.0%]	NCT00374907, NCT00097084	2	480	240
[8.0%,11.0%]	NCT00097084	1	324	324
[11.0%,14.0%]		0	0	0

domain (i.e., “Disease or Syndrome,” “Finding,” etc.) [39] and appears in more than 3% of the studies of a given condition. We excluded the tags of the semantic type “Body Part, Organ, or Organ Component,” because they were not informative in this context. We also divided the features into two groups: qualitative (e.g., diagnosis, procedure, or device) and quantitative (e.g., clinical attribute or laboratory test result). Quantitative features were further parsed as described below.

#### 2.1.4. Extracting quantitative features

We used a numeric expression extraction tool called Valx [48] to structure numeric expressions from free-text eligibility criteria. To identify the quantitative features and extract numeric expressions, Valx pre-compiled a list of unique quantitative features using domain knowledge (from online resources such as WebMD [51], WHO website [52], and web communities for various conditions) and frequent UMLS (Unified Medical Language System) [53] concepts with their synonyms, which cover selected semantic types relevant to quantifiable attributes in clinical studies (e.g., *Clinical Attribute*, *Laboratory or Test Result*, and *Quantitative Concept*). For each quantitative feature, this pre-compiled list contains

possible feature names for unification (e.g., HbA1c, A1c, and Glycohemoglobin), a preferred feature name, allowable measurement units, exception units, a preferred unit for normalization, and a normal range (with a *max\_allowed\_value* and a *min\_allowed\_value*). With the pre-compiled list, Valx unified different feature names and normalized heterogeneous measurement units of the same feature. For example, “A1c,” “hemoglobin A1c,” and “HbA1c” were all recognized as “HbA1c.”

Valx first extracted candidate numeric expressions from eligibility criteria text. Then, using regular expression matching, it structured each candidate numeric expression into four components: a candidate quantitative feature, a comparison symbol (i.e., “=,” “>,” “<,” “>=,” and “<=”), a threshold value, and a measurement unit (e.g., “lb” or “kg”). For example, from an exclusion criterion “evidence for Type 2 diabetes, including fasting plasma glucose greater than or equal to 126 mg/dl or HbA1c greater than 6.5%,” we extracted two numeric expressions: [“Glucose”, “>=”, 126.0, “mg/dL”] and [“HbA1c”, “>”, 6.5, “%”]. Valx built n-grams (i.e., continuous subsequence of n words) of the candidate quantitative feature after removing special characters and punctuation from it. It removed n-grams composed of only English stop-words or irrele-

vant grammatical structures. Each n-gram was matched against the pre-compiled list of quantitative features and retained if at least one substring was in the list. It normalized allowable measurement units to enable meaningful aggregation of quantitative values. Lastly, it performed heuristic rule-based removal of invalid numeric expressions by comparing the threshold value with the allowable value range of the quantitative feature specified in the pre-compiled list, i.e., the expressions with a threshold value above  $max\_allowed\_value * 2$  or below  $min\_allowed\_value/2$  were considered invalid and therefore removed. To unifying the exclusion criteria and inclusion criteria, it converted all numeric expressions in exclusion criteria to inclusion criteria using negations, such as replacing “<”, “<=”, “>”, “>=” with “>=”, “>”, “<=”, “<”, respectively. Valx converted the aforementioned two expressions extracted from the example exclusion criterion to [“Glucose”, “<”, 126.0, “mg/dL”] and [“HbA1c”, “<=”, 6.5, “%”], respectively.

In a previously conducted evaluation of Valx for a paper under preparation, the precision, recall, and F-measure for extracting numeric expressions with the quantitative feature “HbA1c” were 99.6%, 98.1%, 98.8% for Type 1 diabetes trials, and 98.8%, 96.9%, 97.8% for Type 2 diabetes studies, respectively. The results of the corresponding measures for extracting numeric expressions with the quantitative feature “Glucose” were 97.3%, 94.8%, 96.1% for Type 1 diabetes studies, and 92.3%, 92.3%, 92.3% for Type 2 diabetes trials, respectively. We stored in COMPACT all the structured numeric expressions extracted from eligibility criteria of all the downloaded study summaries. These numeric expressions were used in the eligibility feature frequency analysis module and the distribution analysis module (described below).

## 2.2. The eligibility feature frequency analysis module

To provide the user with the list of frequently used eligibility features in a condition-specific set of studies we built an eligibility feature frequency analysis module. Given a condition and the number of frequent features to retrieve – “*K*”, this module will analyze the *K* most frequent qualitative and quantitative eligibility features used by the studies on the selected condition. Multiple occurrences of the same feature in a study are counted only once when computing the frequency. Given this information, when designing eligibility criteria for a new study, a user might reuse some of those frequent features or verify whether a feature s/he considers important is missing.

## 2.3. The query builder

The query builder module allows users to build a distribution analysis query on a selected set of studies. After a user selects a frequent quantitative feature of a condition-specific set of studies, the query builder module will analyze (1) the distribution of number of studies over the value spectrum of the quantitative feature and (2) the distribution of studies by the values of a specific study descriptor (e.g., study phase, intervention type). The distribution of number of studies over the value spectrum of the chosen quantitative feature allows the user to specify a narrower value range of interest for closer investigation. In the query builder, the user must specify (1) a value range (with an upper bound and a lower bound) of the quantitative feature to be analyzed, (2) whether the distribution is on unbinned consecutive value points or binned non-overlapping value intervals, (3) whether the width of value intervals is fixed or varying.

Study descriptors are organized in categories, such as *gender*, *study type*, *intervention type*, *status*, *sponsor type*, and *phase*. Each descriptor value is followed by the percentage of selected studies having that descriptor, which allows the user to estimate the number of studies to be included in the analysis. For example, if a user

chooses “HbA1c” for Type 2 diabetes studies, s/he will see that 95.66% of these studies using “HbA1c” are interventional, whereas 4.29% are observational. A range of age and study start date can be specified to further restrict the list of studies to analyze. The NCT number of studies satisfying the query will be passed to the distribution analysis module for further processing.

## 2.4. The distribution analysis module

Five types of distribution analyses are supported for each quantitative feature as described more fully below: (1) distribution of number of studies over its unbinned consecutive value points or non-overlapping value intervals within the user-specified value range; (2) distribution of enrollment over its value points or value intervals within the user-specified value range; (3) distribution of number of studies over boundary values; (4) distribution of number of studies over permissible value ranges; and (5) distribution of number of studies over permissible value range widths. In the system, we grouped the analysis results of (3–5) together. The user must choose one kind of distribution analysis at a time.

### 2.4.1. Distribution of number of studies

Distribution of number of studies reveals which parts of a quantitative feature’s permissible value range is widely or rarely permitted across the selected studies. This module calculates the distribution of the number of studies over consecutive value points or non-overlapping value intervals of a quantitative feature within the user-specified value range. If a user chooses to plot the distribution of studies on unbinned value points, for each consecutive value in the user-specified value range, we count the number of studies for which patients having the value are eligible. Of note, the distance between two consecutive value points can be fixed (i.e., the user-specified width) or varying (i.e., the actual boundary values are the value points). If s/he chooses to plot on value intervals, we count the number of studies for each value interval within the user-specified value range. One study may cover multiple value points or value intervals. If a user has specified a fixed-width for the value intervals and has picked multiple phases in the query builder, s/he can stratify the distribution by multiple phases in the same figure for comparison. If a user does not pick any phase, studies of any phase (including those that do not specify a phase) will be included in the analysis.

### 2.4.2. Distribution of enrollment

Distribution of enrollment presents the size of the target population across the selected studies for value points or value intervals of the quantitative feature. (In this paper, we do not distinguish between actual and anticipated enrollment due to a data issue in ClinicalTrials.gov. We will discuss this limitation later in Section 4.) If a user chooses to plot the distribution of enrollment on consecutive value points, we sum the enrollment for the studies for which patients having a specific value within the user-specified value range are eligible to obtain the maximum enrollment. If s/he chooses to plot by value intervals, for each interval in the user-specified value range, we sum the enrollment for the studies in the interval to obtain the maximum enrollment. To obtain the average enrollment, for each value point or value interval, we divide the corresponding maximum enrollment by the number of studies. Distribution of enrollment can also be stratified by phases.

### 2.4.3. Distribution of number of studies over boundary values

The distribution of number of studies over boundary values shows the most commonly used upper bound and lower bound of a quantitative feature in selected studies. For this analysis, we retrieve all the boundary values of the feature in the selected studies and count the number of studies for each boundary value. Then

we analyze the distribution stratified by upper bound values and lower bound values. For open value range such as HbA1c  $\geq$  7%, we only include one threshold value 7% in this analysis.

#### 2.4.4. Distribution of number of studies over permissible value ranges

The distribution of the number of studies over permissible value ranges highlights frequently used value ranges of the quantitative feature in the selected studies. We itemize all the permissible value ranges of the chosen quantitative feature in the selected studies, count the number of studies that use each value range, and order the result by the number of studies. In the case of open ranges, we use  $-\infty$  or  $+\infty$  as needed for the missing boundary value. For example, BMI  $\geq$  15 kg/m<sup>2</sup> is denoted as “[15,  $+\infty$ )”, whereas HbA1c < 7.0% is denoted as “( $-\infty$ , 7.0)”. “[ ]” refers to being inclusive, while “( )” refers to being non-inclusive. Of note, in this analysis we do not consider studies that allow only a single value, e.g., the study NCT00519857 with an inclusion criterion “subjects were to have a Body Mass Index (BMI) = 34 kg/m<sup>2</sup>.”

#### 2.4.5. Distribution of number of studies over permissible value range widths

The value range width is the difference between the upper bound and the lower bound of the selected quantitative feature in a study. For example, the value range width of BMI derived from the criterion “BMI value between 20 and 25 kg/m<sup>2</sup>” is 5 (=25 – 20). This distribution reveals frequently used value range width. In this analysis, we do not consider open value ranges such as HbA1c  $\geq$  7.0%, because the difference of  $\infty$  and 7.0 is not a definite number.

### 2.5. Visualization module

The distribution analysis module passes the results of the distribution analysis module to the visualization module. With Google Charts API [54], we plot as line graphs the results of four out of the five analyses described in the previous section: (1) distribution of number of studies (Section 2.4.1), (2) distribution of maximum and average enrollment (Section 2.4.2), (3) distribution of number of studies over boundary values (Section 2.4.3), and (4) distribution of number of studies over permissible value range widths (Section 2.4.5). The line graphs are employed to visualize these four analyses because (1) the  $x$ -axis values are non-overlapping, (2) multiple curves can be drawn on the same line graph (e.g., to show distribution of enrollment stratified by study phases, or distribution of boundary values stratified by upper bound values and lower bound values), and (3) among all the figure types supported by Google Chart API, the line graph is the most suitable figure type to plot two-dimensional distributions of a continuous variable. In contrast, because permissible value ranges of the different studies are likely to be overlapping, we plot the distribution of number of studies over permissible value ranges (Section 2.4.4) as a bar graph, where the value ranges are displayed in descending order of frequency. We leverage the applicable functions of the Google Charts API to make the graphs interactive. A user may see the detail of the analysis result by hovering the mouse cursor on the figure. Multiple curves on the same figure are distinguished by different colors. The user may hide a curve by clicking on its symbol in the legend. The ranges of the  $x$ - and  $y$ -axes of the line graphs can scale itself automatically to the currently visible curves. The user may zoom in on a figure to investigate an area more closely or zoom out to view the global pattern.

### 2.6. The Web-based prototype of VITTA

VITTA is a Web-based implementation of the aforementioned methodology framework. We have made VITTA publicly available

at <http://is.gd/VITTA>. It has five pages for users to interact with: i.e., (1) “condition and feature selection”, (2) “query builder”, and three pages for visualization of the analysis results: (3) distribution of number of studies, (4) distribution of enrollment, and (5) distributions of number of studies over boundary values, permissible value ranges and value range widths.

The “condition and feature selection page” has a dropdown menu containing 1311 conditions pre-indexed by ClinicalTrials.gov. Currently, a user can choose only one value each time. A user can type the first letter of the condition for quick navigation. After a user chooses a condition and specifies the number of frequent features to return, VITTA will analyze qualitative and quantitative eligibility features of the studies indexed to that condition and display a certain number of them (specified by the user) in separate sections in descending order of frequency. The qualitative features are listed in a two-column table (one for inclusion, the other for exclusion criteria) on the web page. Quantitative features are listed in a dropdown menu that allows selection. Currently, our analysis does not include a detailed view of qualitative features beyond their frequency. After the user selects a quantitative feature, VITTA will analyze the characteristics of studies using the quantitative feature and send the user to the “query builder page,” where the user can narrow his/her selection based on values of study descriptors such as study type (interventional or observational), intervention type, study status, sponsor type, and phase. After the user chooses a certain type of distribution analysis and submit the query, s/he can view the analysis results online or download a CSV file that includes NCT numbers and the data for visualizing the results for offline analysis using MATLAB, R, etc. On the result page, the user can also modify the query or start again.

### 2.7. Preliminary evaluation by potential users

In order to assess the user-perceived usefulness of VITTA, we designed a questionnaire consisting of five Likert-scaled [55] multiple-choice questions and four open-ended questions. The five multiple-choice questions adapted the well-adopted System Usability Scale (SUS) [56,57], as shown in Table 3. Of the four open-ended questions, two are about their current practice and the other two are about VITTA. We recruited a convenience sample of five potential users to perform a preliminary evaluation of VITTA. These recruited users had a variety of backgrounds, including three researchers from academia with specialties such as pediatric, cardiology, and health IT usability evaluation as well as two medical informaticians from pharmaceutical companies, and various levels of involvement in clinical research. We conducted the evaluation sessions either in the office of the user or remotely using WebEx. We first gave a comprehensive demonstration of VITTA to the users. Then the users tried the tool with their own queries of interest.

## 3. Results

We downloaded all 163,285 XML files of clinical study summaries (one XML file for each study) registered on ClinicalTrials.gov as of March 18, 2014. After excluding studies without eligibility criteria, 162,586 records were further processed. Using Valx, we extracted 97,560 distinct quantitative features from 682,718 numeric expressions from inclusion criteria and 89,555 distinct quantitative features from 385,421 numeric expressions from exclusion criteria. After unifying exclusion criteria and inclusion criteria, there were in total 180,458 distinct quantitative features, out of which 133 had each a frequency of more than 500.

**Table 3**  
The usefulness questionnaire for VITTA.

Question type	Question number	Question
Multiple-choice	1	When designing new studies, using VITTA may help me identify frequent value ranges of a quantitative feature for eligibility determination
	2	Using VITTA could help explain my rationale for subject selection more explicitly when designing new studies
	3	I find it easier to see global patterns of clinical study eligibility criteria design using VITTA
	4	Assuming VITTA would be available on my job as a clinical trialist, I would use it in the future
	5	I would prefer using VITTA tool for eligibility criteria design rather than manually find similar studies on ClinicalTrials.gov
Open-ended	6	What is your current process for designing clinical study eligibility criteria?
	7	What is the difficulty when designing clinical study eligibility criteria?
	8	What do you see as potential value of VITTA for supporting clinical study designs?
	9	Do you have any suggestions on improving VITTA?

The parsed data are stored in the COMPACT database, currently hosted as a MySQL database instance on Amazon Rational Database Service, which is part of Amazon Web Services. The content of COMPACT will be updated twice a year to reflect the changes in ClinicalTrials.gov. The access to the COMPACT database is given through the Web-based application VITTA described in this paper.

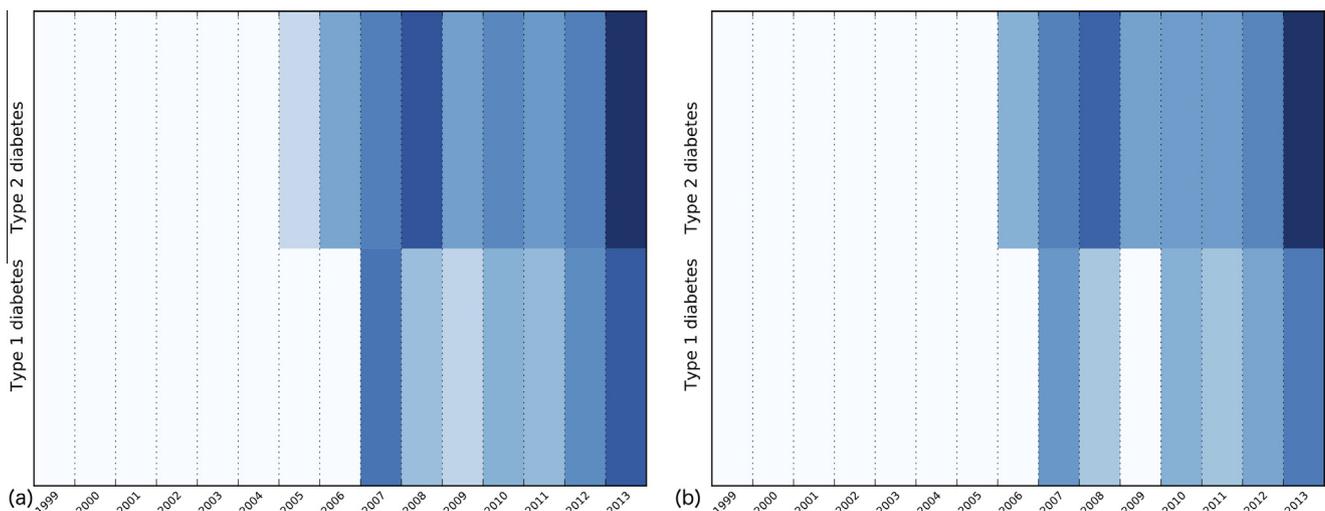
### 3.1. Frequently used eligibility features and trend analysis

On the VITTA interface, users can review a list of the most frequently used qualitative features, where users can select frequently used quantitative features for further distribution analysis. If a user chooses “diabetes mellitus” as the condition of interest, the four most frequently used quantitative features are age (52.89%), HbA1c (36.66%), BMI (32.40%), and Glucose (14.32%). Some quantitative features might have been adopted for participant selection only in recent years. Using a published method for trend analysis of common data elements in clinical study eligibility criteria [41,58], we analyzed the trend over time of quantitative features used by studies of 1311 pre-indexed conditions. Such information can help the designers of a new study see possible changes in participant selection requirements in the studies on a certain condition over a period of time and make informed decisions when designing a new study. Using diabetes as an example, we saw that the use of some quantitative features changed significantly over the past 15 years. Fig. 4 displays the

trend for (a) diastolic blood pressure, and (b) systolic blood pressure. Since 2007, “diastolic blood pressure” and “systolic blood pressure” have been used in eligibility criteria of studies in both types of diabetes. In the future, we will display the trend of a feature on VITTA and allow the users to download the CSV files of the trend analysis of a feature so they will be able to see the number of studies using the feature in each year.

### 3.2. Example distribution analysis queries

Hypertension and Type 2 diabetes are major medical conditions that cause morbidity and mortality worldwide and are therefore given global priorities in clinical research [59]. The eligibility feature frequency analysis module of VITTA identified BMI (Body Mass Index) and HbA1c as representative quantitative features in studies on hypertension and Type 2 diabetes, respectively. This result is in accordance with the literature since BMI has been shown to be closely associated with blood pressure [60,61]. HbA1c test result reflects the average level of blood glucose over the previous three months [62]. Fig. 5 shows the frequent qualitative and quantitative eligibility features if an investigator selects “diabetes mellitus Type 2” as the condition. There are 3897 studies of Type 2 diabetes. HbA1c is used in 48.5% Type 2 diabetes studies, preceded by age with a marginally higher percentage (49.1%). BMI is used in 8.51% hypertension studies, preceded by age and blood pressure measures.



**Fig. 4.** Trend of (a) diastolic blood pressure and (b) systolic blood pressure in diabetes clinical studies from 1999 to 2013. The horizontal line indicates 15 years (1999–2013), while the vertical bars indicate Type 1 or Type 2 diabetes, respectively. Each blue marker indicates that the quantitative feature appears in at least 3% of the studies for the corresponding diabetes type for the corresponding year. The darker the shade of blue, the more frequently the quantitative feature is used in studies of the corresponding year. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

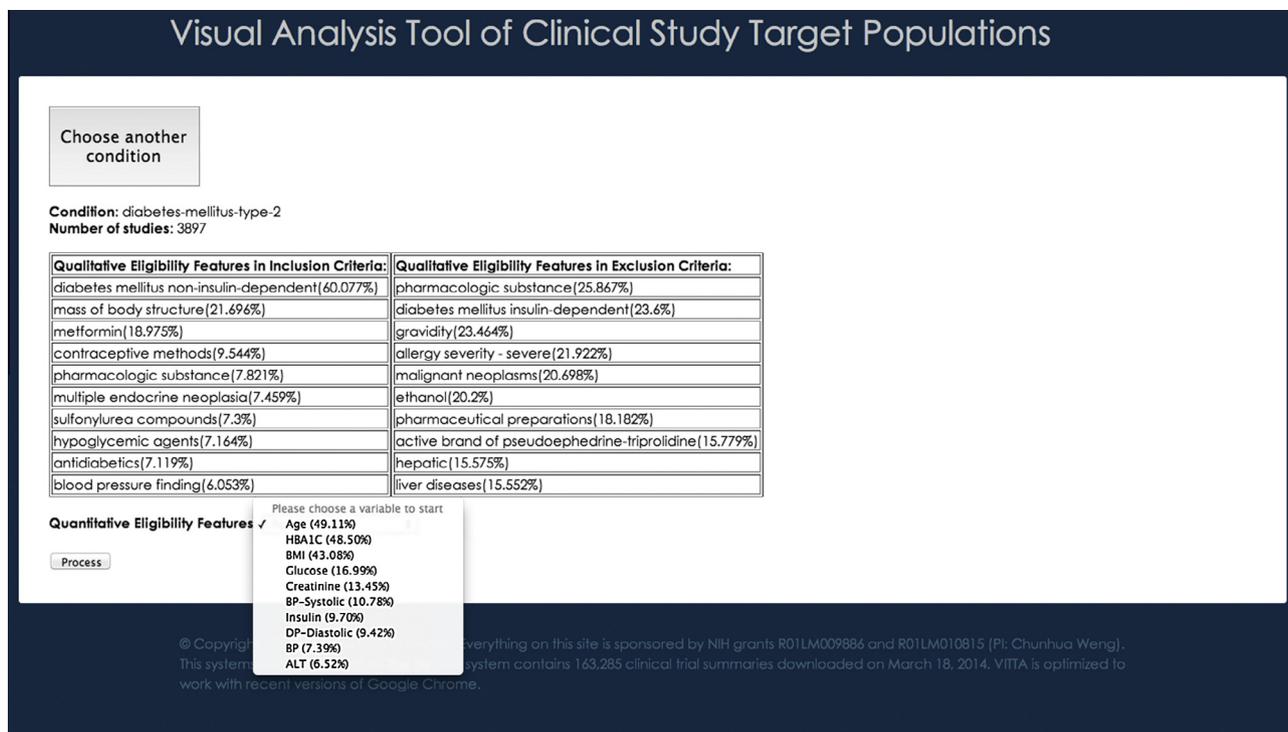


Fig. 5. The frequent qualitative and quantitative features of Type 2 diabetes studies.

### 3.2.1. Query builder: Distribution of studies by selected descriptors

Table 4 shows the distribution by selected study descriptors on the query builder page after an investigator chooses the quantitative feature to be “HbA1c” for Type 2 diabetes studies or “BMI” for hypertension studies. Note that some studies use “Phase 1/Phase 2” or “Phase 2/Phase 3” in the phase field. When an investigator selects “Phase 1”, both “Phase 1” and “Phase 1/Phase 2” studies are included for the distribution analyses.

Next we demonstrate the utility of VITTA using two sample queries, which include the quantitative features “BMI” and “HbA1c” for conditions “hypertension” and “Type 2 diabetes,” respectively. Note that for each query, we only present a subset of the figures that can be generated by VITTA.

### 3.2.2. Query 1: For interventional studies of hypertension using drug(s) as intervention and using BMI as eligibility feature, what is the distribution of number of studies stratified by phases 1/2/3/4 and the distribution of enrollment on fixed-width (0.5 kg/m<sup>2</sup>) value points of BMI between 15 kg/m<sup>2</sup> and 40 kg/m<sup>2</sup>?

There were 164 studies that satisfied the query (condition: hypertension, study type: interventional, intervention type: drug, phases: 1/2/3/4, quantitative eligibility feature: BMI, user-specified value range: [15 kg/m<sup>2</sup>, 40 kg/m<sup>2</sup>]). Fig. 6(a) shows the distribution of number of studies in different phases. The x-axis shows the consecutive value points of BMI of fixed-width (0.5 kg/m<sup>2</sup>), while the y-axis shows the number of studies. Due to limited space, the numeric labels of x-axis are not completely displayed. The distribution of maximum enrollment is shown in Fig. 6(b) where the y-axis shows the maximum enrollment. The user can observe that more participants with BMI value between 25 kg/m<sup>2</sup> and 40 kg/m<sup>2</sup> are enrolled than those with BMI value between 15 kg/m<sup>2</sup> and 25 kg/m<sup>2</sup>. Comparing Fig. 6(a) and (b), we found that even though there are similar numbers of studies recruiting patients with BMI values between 19 kg/m<sup>2</sup> and 25 kg/m<sup>2</sup> and between 25 kg/m<sup>2</sup> and 35 kg/m<sup>2</sup>, the enrollment in studies requiring the first range is noticeably

smaller than that of the latter. This is likely because individuals with the latter range of BMI, which indicates overweight and obesity [63], are at greater risk of hypertension [64]. Fig. 6(c) shows that the average enrollment is gradually increasing with the increasing value range of BMI. On average, phase 3 studies recruit more patients with BMI greater than 27 kg/m<sup>2</sup> than studies in other phases.

### 3.2.3. Query 2: For Type 2 diabetes industry-sponsored studies in phases 1/2/3/4 that use HbA1c as eligibility feature and are recruiting, what is the distribution of maximum enrollment for HbA1c of fixed-width (0.5%) value intervals of HbA1c between 1% and 12%? What is the distribution of number of studies over boundary values, permissible value ranges, and permissible value range widths?

For this query, VITTA analyzed 113 qualifying studies (condition: diabetes mellitus Type 2, status: recruiting, phases: 1/2/3/4, sponsor type: industry, quantitative feature: HbA1c, user-specified value range: [1%, 12%]). The distribution of maximum enrollment is displayed in Fig. 7(a). We can see that most Type 2 diabetes industry-sponsored studies recruit patients with HbA1c values between 7.0% and 10.0%, which conforms to the finding in clinical practice: HbA1c levels higher than 6.5% indicate diabetes [65]. Fig. 7(b) shows the distribution of number of studies over boundary values. To better display the modal upper and lower boundary values, we hid the bimodal curve for the overall boundary values and left two separated curves for the values used as upper bound or lower bound, respectively. We can see that 7.0% is the most frequently used lower bound, whereas 10.0% is the most frequently used upper bound. Fig. 7(c) shows that the most used permissible value range of HbA1c in the selected studies are [7.5,9.0] and [7.0,9.5]. Fig. 7(d) shows that “1.5” and “2.5” are the most used value range widths in the selected studies that specify both an upper bound and a lower bound of HbA1c.

**Table 4**  
Distribution by selected descriptors of Type 2 diabetes studies using HbA1c and hypertension studies using BMI for participant selection.

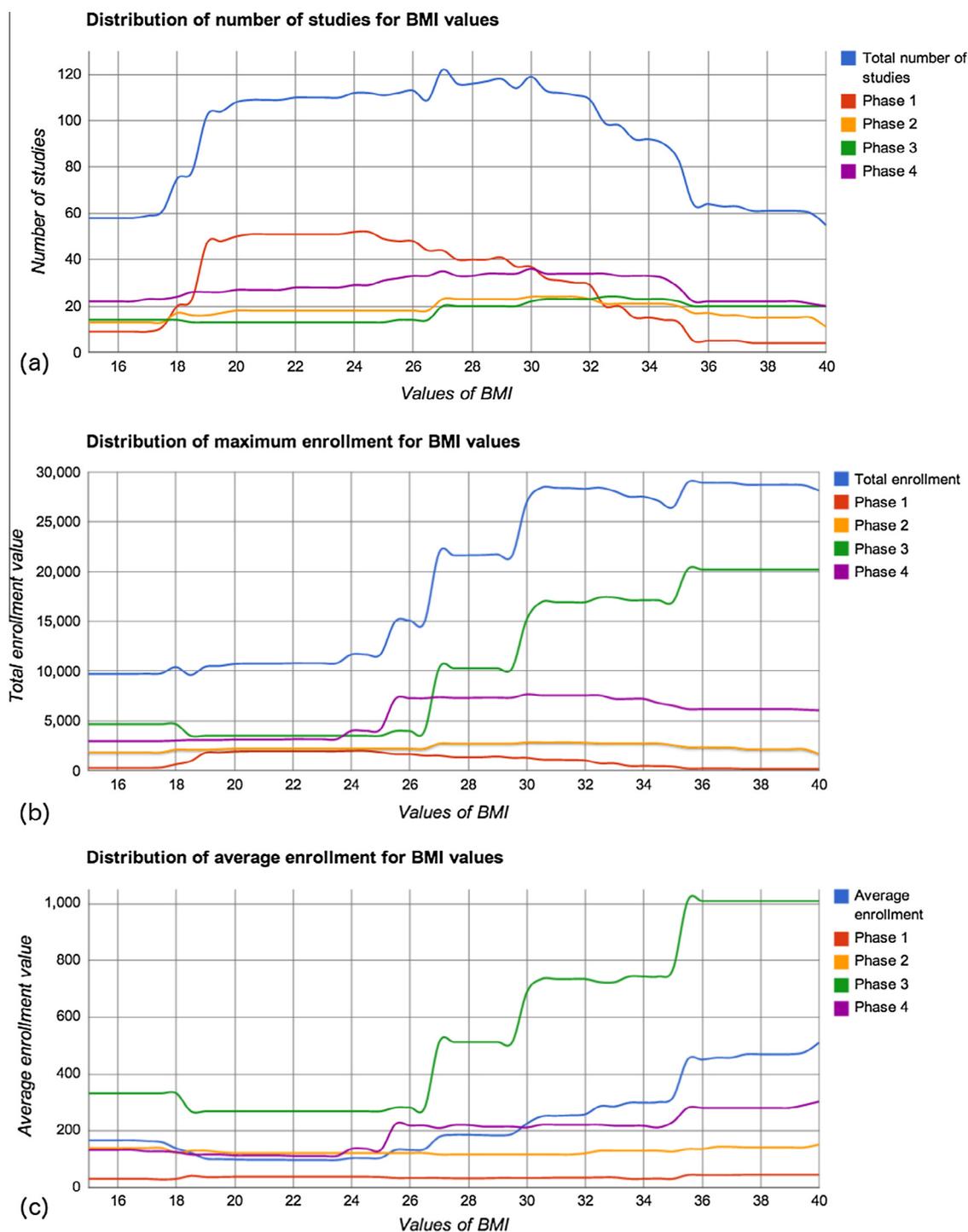
Study characteristics	Quantitative features	
	HbA1c (Type 2 diabetes) (n = 1890)	BMI (hypertension) (n = 389)
<i>Study type</i>	1890	389
Interventional	1808 (95.66%)	356 (91.52%)
Observational	81 (4.29%)	33 (8.48%)
Observational [Patient Registry]	1 (0.05%)	0 (0%)
Expanded access	0 (0%)	0 (0%)
<i>Intervention type</i>	1833	362
Drug	1423 (77.63%)	212 (58.56%)
Procedure	51 (2.78%)	14 (3.87%)
Biological	20 (1.09%)	4 (1.10%)
Device	60 (3.27%)	10 (2.76%)
Behavioral	117 (6.38%)	58 (16.02%)
Dietary supplement	69 (3.76%)	35 (9.67%)
Genetic	8 (0.05%)	0 (0%)
Radiation	1 (0.05%)	0 (0%)
Other	92 (5.02%)	29 (8.01%)
<i>Status</i>	1890	389
Open studies	442 (23.39%)	105 (26.99%)
Recruiting	367 (19.42%)	81 (20.82%)
Not yet recruiting	75 (3.97%)	24 (6.17%)
Closed studies	1448 (76.61%)	284 (73.01%)
Active, not recruiting	135 (7.14%)	46 (11.83%)
Completed	1179 (62.38%)	210 (53.98%)
Withdrawn	18 (0.95%)	3 (0.77%)
Suspended	11 (0.58%)	1 (0.26%)
Terminated	95 (5.03%)	22 (5.66%)
Enrolling by invitation	10 (0.53%)	2 (0.51%)
<i>Sponsor type</i>	1890	389
NIH	19 (1.01%)	14 (3.60%)
Industry	1159 (61.32%)	135 (34.70%)
Other U.S. Federal Agency	13 (0.69%)	6 (1.54%)
Other	699 (36.98%)	234 (60.15%)
<i>Phase</i>	1890	389
0	9 (0.48%)	4 (1.03%)
1	166 (8.78%)	69 (17.74%)
2	292 (15.45%)	39 (10.03%)
3	539 (28.52%)	42 (10.80%)
4	400 (21.16%)	54 (13.88%)
Not applicable or not specified	408 (21.59%)	165 (42.42%)
<i>Gender</i>	1890	389
Both	1827 (96.67%)	330 (84.83%)
Male only	39 (2.06%)	22 (5.66%)
Female only	24 (1.27%)	37 (9.51%)
<i>Study design – intervention model (for interventional studies)</i>	1890	389
Parallel	1364 (72.17%)	199 (51.16%)
Factorial	28 (1.48%)	7 (1.80%)
Crossover	218 (11.53%)	97 (24.94%)
Single group	186 (9.84%)	53 (13.62%)
Not applicable or not specified	94 (4.97%)	33 (8.48%)
<i>Study design – allocation (for interventional studies)</i>	1890	389
Randomized	1595 (84.39%)	306 (78.66%)
Non-randomized	113 (5.98%)	19 (4.88%)
Not applicable or not specified	181 (9.58%)	64 (16.45%)
<i>Study design – time perspective (for observational studies)</i>	1890	389
Cross-sectional	12 (0.63%)	4 (1.03%)
Longitudinal	1 (0.05%)	1 (0.26%)
Prospective	64 (3.39%)	19 (4.88%)
Retrospective	64 (0.16%)	3 (0.77%)
Not applicable or not specified	1810 (95.77%)	362 (93.06%)

### 3.3. Results of preliminary evaluation with five potential users

Table 5 lists the results of the multiple-choice questions (Q1–Q5) answered by five potential users in the preliminary evaluation. The SUS-based score  $(4.6 + 4.2 + 4.0 + 4.2 + 4.6) * 4 = 86.4$  suggests the usefulness of VITTA for researchers involved in the design and evaluation of clinical studies. It also confirmed that this method has the potential to help profile aggregated target popula-

tions for sets of studies and enable the generalizability analysis of those studies.

Most of the five recruited potential users expressed keen interest in using the tool for investigating eligibility criteria of existing studies or for designing new studies. Regarding the current process of designing clinical study eligibility criteria, User A “usually uses literature search to find similar studies.” User B said: “we always make a team decision with domain experts,” and he usually “looks



**Fig. 6.** For consecutive BMI values, (a) distribution of studies in phases 1/2/3/4; (b) distribution of maximum enrollment for interventional drug studies on hypertension in phases 1/2/3/4; (c) distribution of average enrollment for interventional drug studies on hypertension in phases 1/2/3/4.

at previous study protocols.” User C “uses criteria from established studies within the organization.” Regarding the difficulty encountered when designing clinical study eligibility criteria, User A pointed out that “it is impossible to look at studies one by one to design eligibility criteria.” User B finds it difficult to “assess the protocol feasibility in terms of how many patients can be recruited from the real world.” User C pointed at the “lack of uniformity of actual criteria. There are multiple ways to say the same thing.” User D “finds it difficult to find similar clinical studies.” Regarding the potential

value of VITTA, they noted that they would be able to learn from large number of existing studies, which can save them a lot of time.

The five potential users collectively suggested that VITTA could be used in the following scenarios: (1) to find commonly used permissible value range(s) of a quantitative eligibility feature, (2) to find potentially under-represented target populations for clinical studies, and (3) to support the rationale for using a certain value range for grant applications. In terms of improvements for VITTA, User C suggested the use of specific eligibility features, for exam-

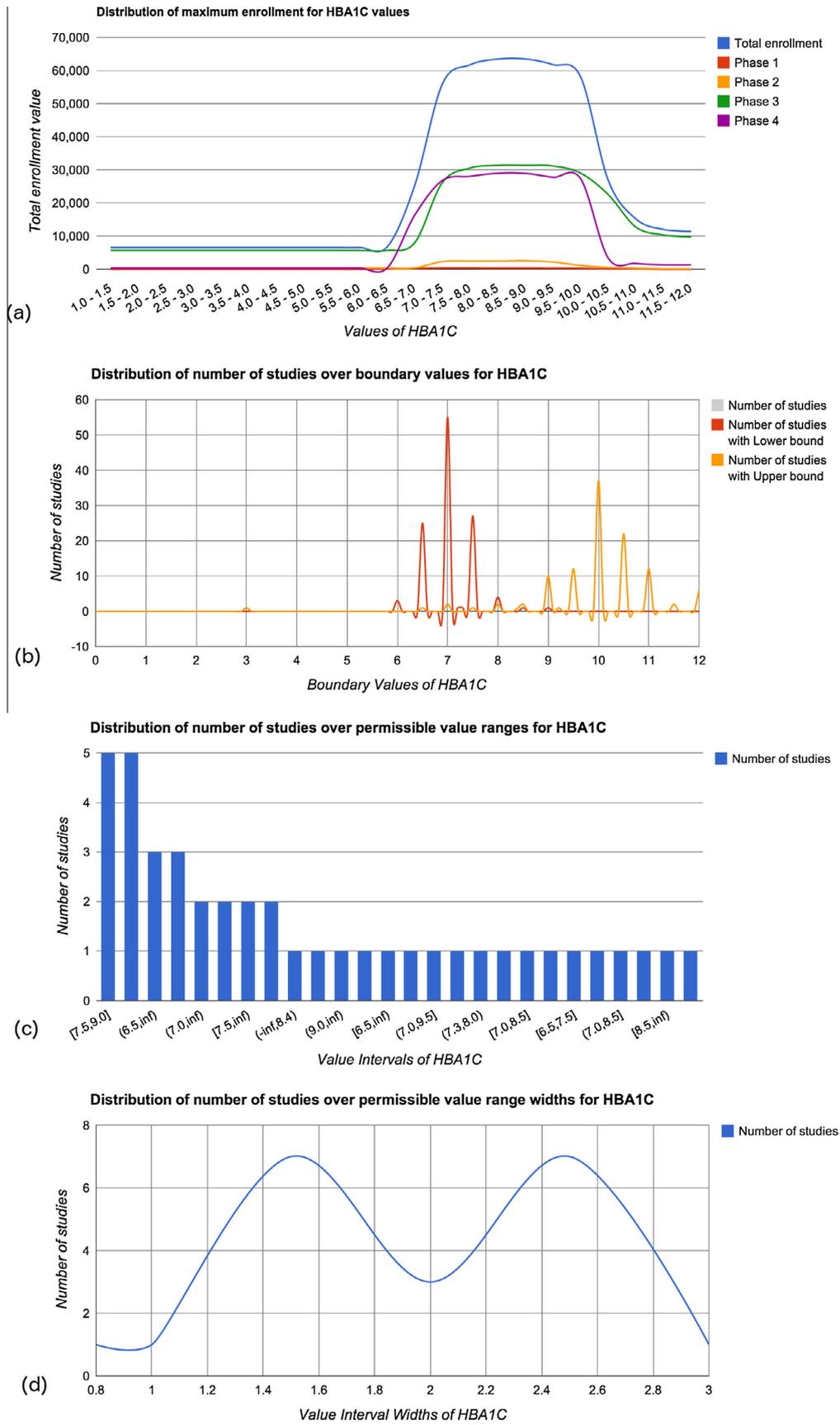


Fig. 7. In industry-sponsored Type 2 diabetes studies in phases 1/2/3/4 that are recruiting, (a) distribution of maximum enrollment over HbA1c's value intervals; (b) distribution of number of studies over boundary values of HbA1c; (c) distribution of number of studies over permissible value ranges of HbA1c; (d) distribution of number of studies over permissible value range widths of HbA1c.

**Table 5**

The usefulness evaluation results (see Table 3 for the questions).

Question number	Average score	Strongly agree (5) (%)	Agree (4) (%)	Neutral (3)	Disagree (2)	Strongly disagree (1)
1	4.6	60	40			
2	4.2	40	40	20%		
3	4.0	40	40		20%	
4	4.2	60	20		20%	
5	4.6	60	40			

ple, “pharmacologic substance” should be replaced with more specific substance concepts. User E suggested aligning the workflow with the development process of clinical studies.

#### 4. Discussion

Our results confirmed the feasibility and usefulness of profiling the quantitative eligibility features of the target populations of sets of clinical studies. When complemented by similar work on qualitative features, our approach promises to enhance the ability to characterize, understand, and improve clinical research generalizability. Our preliminary evaluation suggests that our method has the potential to improve the accessibility of design patterns in eligibility criteria through a global view of sets of clinical studies with common characteristics. It is laborious and sometimes impossible to obtain such a global view through manual review of many clinical study summaries.

Our method also establishes a highly desirable prerequisite for generalizability analyses of sets of clinical studies, which involve comparisons between target populations and real-world patient populations. A data-driven analytical system such as VITTA profiles the distribution of the collective target populations in multiple existing studies along the value spectrum of a selected quantitative eligibility feature, and can potentially guide the participant selection of future studies in an interactive fashion. Stakeholders of clinical trials can use VITTA to detect potential hidden eligibility criteria design biases towards certain population subgroups at the clinical research community level. For example, if most Type 2 diabetes mellitus trials require eligible patients to have their A1c equal to or above 7%, investigation may be warranted to justify why those patients whose A1c values are between 6.5% and 7% are not considered for these trials. In the clinical trials research community, thresholds for blood pressure (e.g., 120 mmHg vs. 140 mmHg) for defining hypertension or thresholds for A1c for defining pre-diabetic and diabetic population are still open research topics. Tools like VITTA can effectively make transparent the collective design trends among clinical trial designers. Although we chose only two common conditions and their representative quantitative features to illustrate our analytical framework, VITTA is designed to support distribution analyses of any quantitative feature of selected studies of any condition assuming that the names and measurement units of all corresponding numeric expressions are correctly recognized and parsed. The modular design of VITTA enables flexible customization of each module for different tasks. For example, the core of the visualization module can be replaced by D3 JavaScript library with more interactive functionalities than Google Charts API. As summary results of more and more studies are added to ClinicalTrials.gov [66], we imagine that systematic reviewers could use VITTA to identify studies of interest and to profile their target populations or study populations.

##### 4.1. Limitations

Quite a number of limitations are noteworthy in interpreting this study. One caveat is that the current version of the quantita-

tive feature extraction tool Valx [48] employed by this study needs more formal evaluation and further improvement. Currently, even though we have filtered out many meaningless tags using pre-defined heuristics, the complexity and heterogeneity of free-text eligibility criteria makes it a daunting task to unify all the quantitative eligibility feature names and all the different measurement units of a quantitative feature. In the future, we need more scalable solutions to improve the accuracy of numeric expression parsing for other quantitative features. In spite of the need to improve the accuracy of the natural language processing (NLP) techniques employed, instead of waiting for perfect NLP techniques, we presented this novel use of “good-enough” NLP results of clinical trial summaries to motivate more NLP researchers to join our efforts in extracting patterns from eligibility criteria text. The modular design of VITTA makes it easy to provide more accurate distribution analysis results when more accurate parsing results are updated in the COMPACT database.

Another limitation is that our method was designed specifically for ClinicalTrials.gov, which does not include all studies worldwide. Studies that are not subject to US law may be registered only in the relevant countries. Many international studies registered in their corresponding countries are included in the International Clinical Trials Registry Platform (ICTRP) [67] managed by the World Health Organization. Nevertheless, about 80% of studies in ICTRP were also registered in ClinicalTrials.gov [23], indicating that our analytical framework includes the majority of clinical study summaries available on the Internet.

Many studies may have partial or condensed eligibility criteria entered in ClinicalTrials.gov [68]. For some conditions, there might not be a representative quantitative feature to profile clinical study target populations. When indexing studies for a condition, the studies returned from ClinicalTrials.gov might include studies of a similar but not the same condition. For example, a search on condition “hypertension” will also retrieve studies on “pulmonary hypertension” which is a different condition from hypertension (high blood pressure). According to COMPACT, about 40% studies in ClinicalTrials.gov did not specify their study phases. When a user selects studies of a certain phase, some studies might be left out because the phase information is not available or applicable. When interpreting the distribution analysis results generated by VITTA, the limitations in ClinicalTrials.gov itself should be taken into account.

The last major limitation is that ClinicalTrials.gov defines only one field for “enrollment”, which is initially used to indicate anticipated enrollment before study starts but later used to indicate the actual enrollment after the completion of the study. Therefore, the meaning of this data field is not stable throughout the study. We suggest that both anticipated enrollment and actual enrollment be reported separately for comparing the difference between target populations and actually enrolled patient populations. This is a design that can be incorporated to improve ClinicalTrials.gov.

##### 4.2. Ongoing and future work

Our future work along this line will focus on adding more study descriptors of VITTA for selecting studies, e.g., interventions, spon-

sors, locations, etc. Currently, it only stratifies the distribution of studies or enrollment by phase. In the future, it will support study stratification using other study descriptors, e.g., study design, recruiting status, etc. To test the feasibility of this system, we recruited only a convenience sample of a small number of potential users to evaluate the usefulness of VITTA. In the future, a larger and more comprehensive evaluation with clinical investigators and sponsors from both academia and industry can provide more insights into their needs for the analysis of clinical study target populations. We also acknowledge the limitation in profiling target populations using only one quantitative feature each time. Therefore, we are investigating analytical methods for profiling target populations using multiple eligibility features simultaneously. After incorporating the data of real patients, we also plan to analyze how the application order of eligibility criteria would affect the cohort size, which might provide actionable knowledge to accelerate patient recruitment.

## 5. Conclusions

We contribute a novel method for profiling the collective target populations of sets of studies registered in ClinicalTrials.gov based on quantitative eligibility features extracted from free-text eligibility criteria. The Web-based system VITTA enables interactive distribution analyses and visualization of frequently used eligibility features. Our preliminary evaluation with a small number of potential users demonstrated the potential of VITTA for improving the transparency for participant selection for clinical research. This method can also potentially help clinical trialists, patients, sponsors, and policy makers identify global patterns in eligibility criteria in studies registered in ClinicalTrials.gov. In the future we will conduct more evaluation studies to assess how VITTA can support patient-centered outcomes research and improve the usability and utility of VITTA based on user feedback accordingly.

## Funding

This study is sponsored by Grants R01LM009886 from the United States National Library of Medicine (PI: Weng) and UL1 TR000040 from the United States National Center for Advancing Translational Sciences (PI: Ginsberg).

## Contributions of authors

All authors meet the ICMJE criteria for authorship. ZH performed system development, analysis and interpretation of data, and drafting of the manuscript. SC and IS participated in the system design and edited the manuscript critically. CW conceptualized and directed the research and made significant methodology contributions and edits to this paper.

## Ethics approval

The Columbia University Medical Center Institutional Review Board approved the study.

## Acknowledgments

We thank Dr. Tianyong Hao for providing Valx for extracting and parsing numeric expressions from free-text eligibility criteria, Dr. Riccardo Miotto for helping us mine frequent qualitative eligibility features and for analyzing the trends of quantitative eligibility features, and potential users, Drs. Melissa Stockwell, Carlos Lopez-Jimenez, John Cai, Michael Cantor, and Po-Yin Yen for providing feedback regarding the value and usability of VITTA.

## References

- [1] From the NIH director: the importance of clinical trials. <<http://www.nlm.nih.gov/medlineplus/magazine/issues/summer11/articles/summer11pg2-3.html>> [09.04.14].
- [2] Somerson JS, Bhandari M, Vaughan CT, Smith CS, Zelle BA. Lack of diversity in orthopaedic trials conducted in the United States. *J Bone Joint Surg Am* 2014;96(7):e56.
- [3] Rodriguez MI, Gordon-Maclean C. The safety, efficacy and acceptability of task sharing tubal sterilization to midlevel providers: a systematic review. *Contraception* 2014;89(6):504–11.
- [4] Blanco C, Olfson M, Goodwin RD, Ogburn E, Liebowitz MR, Nunes EV, et al. Generalizability of clinical trial results for major depression to community samples: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *J Clin Psychiatr* 2008;69(8):1276–80.
- [5] Hamann GF. Early ischemic signs should not be used as exclusion criteria in thrombolysis trials. *Stroke* 2004;35(1):e3–4.
- [6] Hoertel N, Le Strat Y, De Maricourt P, Limosin F, Dubertret C. Are subjects in treatment trials of panic disorder representative of patients in routine clinical practice? Results from a national sample. *J Affect Disorders* 2013;146(3):383–9.
- [7] Hoertel N, Le Strat Y, Lavaud P, Dubertret C, Limosin F. Generalizability of clinical trial results for bipolar disorder to community samples: findings from the National Epidemiologic Survey on Alcohol and Related Conditions. *J Clin Psychiatr* 2013;74(3):265–70.
- [8] Hoertel N, Le Strat Y, Limosin F, Dubertret C, Gorwood P. Prevalence of subthreshold hypomania and impact on internal validity of RCTs for major depressive disorder: results from a national epidemiological sample. *PLoS One* 2013;8(2):e55448.
- [9] Humphreys K, Maisel NC, Blodgett JC, Finney JW. Representativeness of patients enrolled in influential clinical trials: a comparison of substance dependence with other medical disorders. *J Stud Alcohol Drugs* 2013;74(6):889–93.
- [10] Okuda M, Hasin DS, Olfson M, Khan SS, Nunes EV, Montoya I, et al. Generalizability of clinical trials for cannabis dependence to community samples. *Drug Alcohol Depen* 2010;111(1–2):177–81.
- [11] Ronconi JM, Shiner B, Watts BV. Inclusion and exclusion criteria in randomized controlled trials of psychotherapy for PTSD. *J Psychiatr Pract* 2014;20(1):25–37.
- [12] Schmidt AF, Groenwold RHH, van Delden JJM, van der Does Y, Klungel OH, Roes KCB, et al. Justification of exclusion criteria was underreported in a review of cardiovascular trials. *J Clin Epidemiol* 2014;67(6):635–44.
- [13] Schneider LS, Olin JT, Lyness SA, Chui HC. Eligibility of Alzheimer's disease clinic patients for clinical trials. *J Am Geriatr Soc* 1997;45(8):923–8.
- [14] Zimmermann JB, Horsch J, Weigand MA, Bruckner T, Martin EO, Hoppe-Tichy T, et al. Patients enrolled in randomised clinical trials are not representative of critically ill patients in clinical practice. *Observational study focus on tigecycline*. *Int J Antimicrob Agents* 2013;42(5):436–42.
- [15] Zimmerman M, Chelminski I, Posternak MA. Exclusion criteria used in antidepressant efficacy trials: consistency across studies and representativeness of samples included. *J Nerv Ment Dis* 2004;192:87–94.
- [16] Schoenmaker N, Van Gool WA. The age gap between patients in clinical studies and in the general population: a pitfall for dementia research. *Lancet Neurol* 2004;3(10):627–30.
- [17] Hao T, Rusanov A, Boland MR, Weng C. Clustering clinical trials with similar eligibility criteria features. *J Biomed Inform* 2014. pii: S1532-0464(14)00011-2.
- [18] NCATS. NCATS Advisory Council Working Group on the IOM Report: The CTSA Program at NIH. 2014; May 16, 2014.
- [19] Milian K, Bucur A, Harmelen FV. Building a library of eligibility criteria to support design of clinical trials. *Knowl Eng Knowl Manage Lect Notes Comput Sci* 2012;7603:327–36.
- [20] ClinicalTrials.gov. <<http://www.clinicaltrials.gov/>> [February 2014].
- [21] History, Policies, and Laws of ClinicalTrials.gov. <<http://www.clinicaltrials.gov/ct2/about-site/history>> [09.04.14].
- [22] Tasneem A, Aberle L, Ananth H, Chakraborty S, Chiswell K, McCourt BJ, et al. The database for Aggregate Analysis of ClinicalTrials.gov (AACT) and subsequent regrouping by clinical specialty. *PLoS ONE* 2012;7(3):e33677.
- [23] Goswami ND, Pfeiffer CD, Horton JR, Chiswell K, Tasneem A, Tsalik EL. The state of infectious diseases clinical trials: a systematic review of ClinicalTrials.gov. *PLoS One* 2013;8(10):e77086.
- [24] Hirsch BR, Califf RM, Cheng SK, Tasneem A, Horton J, Chiswell K, et al. Characteristics of oncology clinical trials: insights from a systematic analysis of ClinicalTrials.gov. *JAMA. Intern Med* 2013;173(11):972–9.
- [25] Lakey WC, Barnard K, Batch BC, Chiswell K, Tasneem A, Green JB. Are current clinical trials in diabetes addressing important issues in diabetes care? *Diabetologia* 2013;56(6):1226–35.
- [26] Ross J, Tu S, Carini S, Sim I. Analysis of eligibility criteria complexity in clinical trials. *AMIA Summits Transl Sci Proc* 2010;2010:46–50.
- [27] Overby CL, Pathak J, Haerian K, Perotte A, Murphy S, Bruce K, et al. A collaborative approach to develop an electronic health record phenotyping algorithm for drug-induced liver injury. *J Am Med Inform Assoc* 2013;20:e243–52.
- [28] Miotto R, Weng C. Unsupervised mining of frequent tags for clinical eligibility text indexing. *J Biomed Inform* 2013;46(6):1145–51.

- [29] Miotto R, Jiang S, Weng C. EFACTS: a method for dynamically filtering clinical trial search results. *J Biomed Inform* 2013;46(6):1060–7.
- [30] Boland MR, Miotto R, Weng C. A method for probing disease relatedness using common clinical eligibility criteria. *Stud Health Technol Inform* 2013;192:481–5.
- [31] Hao T, Rusanov A, Weng C. Extracting and normalizing temporal expressions in clinical data requests from researchers. In: International conference for smart health, Lecture notes in computer science, vol. 8040; 2013. p. 41–51.
- [32] Boland MR, Miotto R, Gao J, Weng C. Feasibility of feature-based indexing, clustering, and search of clinical trials. A case study of breast cancer trials from ClinicalTrials.gov. *Methods Inf Med* 2013;52(5):382–94.
- [33] Luo Z, Miotto R, Weng C. A human-computer collaborative approach to identifying common data elements in clinical trial eligibility criteria. *J Biomed Inform* 2013;46(1):33–9.
- [34] Weng C, Wu X, Luo Z, Boland MR, Theodoratos D, Johnson SB. EliXR: an approach to eligibility criteria extraction and representation. *J Am Med Inform Assoc* 2011;18(Suppl. 1):i116–24.
- [35] Luo Z, Yetisgen-Yildiz M, Weng C. Dynamic categorization of clinical research eligibility criteria by hierarchical clustering. *J Biomed Inform* 2011;44(6):927–35.
- [36] Luo Z, Johnson SB, Lai AM, Weng C. Extracting temporal constraints from clinical research eligibility criteria using conditional random fields. *AMIA Annu Symp Proc* 2011;2011:843–52.
- [37] Boland MR, Tu SW, Carini S, Sim I, Weng C. EliXR-TIME: a temporal knowledge representation for clinical research eligibility criteria. *AMIA Summits Transl Sci Proc* 2012;2012:71–80.
- [38] Weng C, Tu SW, Sim I, Richesson R. Formal representation of eligibility criteria: a literature review. *J Biomed Inform* 2010;43(3):451–67.
- [39] Luo Z, Johnson SB, Weng C. Semi-automatically inducing semantic classes of clinical research eligibility criteria using UMLS and hierarchical clustering. *AMIA Annu Symp Proc* 2010;2010:487–91.
- [40] Luo Z, Duffy R, Johnson S, Weng C. Corpus-based approach to creating a semantic lexicon for clinical research eligibility criteria from UMLS. *AMIA Summits Transl Sci Proc* 2010;2010:26–30.
- [41] Weng C, Yaman A, Lin K, He Z. Trend and network analysis of common eligibility features for cancer trials in ClinicalTrials.gov. In: International conference for smart health, Lecture notes in computer science, vol. 8549; 2014. p. 130–41.
- [42] He Z, Carini S, Hao T, Sim I, Weng C. A method for analyzing commonalities in clinical trial target populations. *AMIA Annu Symp Proc* 2014;2014:1777–86.
- [43] Weng C, Li Y, Ryan P, Zhang Y, Gao J, Liu F, et al. A distribution-based method for assessing the differences between clinical trial target populations and patient populations in electronic health records. *Appl Clin Inform* 2014;5(2):463–79.
- [44] Abousalh-Neto NA, Kazgan S, editors. Big data exploration through visual analytics. In: 2012 IEEE conference on visual analytics science and technology. Seattle (WA); 2012.
- [45] Simpao AF, Ahumada LM, Galvez JA, Rehman MA. A review of analytics and clinical informatics in health care. *J Med Syst* 2014;38(4):45.
- [46] Gotz D, Wongsuphasawat K. Interactive intervention analysis. *AMIA Annu Symp Proc* 2012;2012:274–80.
- [47] Gotz D, Wang F, Perer A. A methodology for interactive mining and visual analysis of clinical event patterns using electronic health record data. *J Biomed Inform* 2014;48:148–59.
- [48] Hao T, Weng C. Valx – numeric expression extraction and normalization tool; 2013. <<http://columbiaelixer.appspot.com/valx>> [February 2014].
- [49] NLM. List of medical conditions in ClinicalTrials.gov. <[http://www.clinicaltrials.gov/ct2/search/browse?brwse=cond\\_alpha\\_all](http://www.clinicaltrials.gov/ct2/search/browse?brwse=cond_alpha_all)> [March 2014].
- [50] How to find studies by topic or on a map in ClinicalTrials.gov. <<https://clinicaltrials.gov/ct2/help/how-find/browse#Topic>> [August 2014].
- [51] WebMD homepage. <<http://www.webmd.com/>> [May 2014].
- [52] WHO homepage. <<http://www.who.int/>> [May 2014].
- [53] Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucl Acids Res* 2004;32:D267–70 (Database issue).
- [54] Google Charts API homepage. <<https://developers.google.com/chart/>> [07.04.14].
- [55] Jamieson S. Likert scales: how to (ab)use them. *Med Educ* 2004;38(12):1217–8.
- [56] Brooke J. SUS. A 'quick and dirty' usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester BA, editors. Usability evaluation in industry. London: Taylor and Francis; 1996. p. 189–94.
- [57] Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. *J Usability Stud* 2009;4:114–23.
- [58] Weng C. Temporal knowledge acquisition from clinical research documents for community-based clinical research data standards development. In: 2014 AMIA joint summits on translational science. Podium abstract; 2014.
- [59] WHO. Burden: mortality, morbidity and risk factors. Global status report on noncommunicable diseases 2010; 2011.
- [60] Egan BM, Zhao Y, Axon RN. US trends in prevalence, awareness, treatment, and control of hypertension, 1988–2008. *JAMA* 2010;303(20):2043–50.
- [61] Abu-Saad K, Chetrit A, Eilat-Adar S, Alpert G, Atamna A, Gillon-Keren M, et al. Blood pressure level and hypertension awareness and control differ by marital status, sex, and ethnicity: a population-based study. *Am J Hypertens* 2014.
- [62] MedlinePlus. <<http://www.nlm.nih.gov/medlineplus/>> [February 2014].
- [63] NHLBI. Range of body mass index. <<https://http://www.nhlbi.nih.gov/guidelines/obesity/BMI/bmicalc.htm>> [April 2014].
- [64] Foulds HJ, Bredin SS, Warburton DE. The relationship between hypertension and obesity across different ethnicities. *J Hypertens* 2012;20(2):359–67.
- [65] HbA1c value range. <<http://www.webmd.com/diabetes/guide/glycated-hemoglobin-test-hba1c>> [February 2014].
- [66] Hudson KL, Collins FS. Sharing and reporting the results of clinical trials. *JAMA* 2015;313(4):355–6.
- [67] International Clinical Trials Registry Platform (ICTRP) Homepage. <<http://www.who.int/ictpr/en/>> [09.04.14].
- [68] Bhattacharya S, Cantor MN. Analysis of eligibility criteria representation in industry-standard clinical trial protocols. *J Biomed Inform* 2013;46(5):805–13.