Implementation of inclusion and exclusion criteria of clinical studies in the software OHDSI ATLAS

Romina Blasini (romina.blasini@informatik.med.uni-giessen.de)
Institute of Medical Informatics, University of Giessen

Kornelia Marta Buchowicz
Faculty of Health Sciences, University of Applied Sciences

Henning Schneider
Institute of Medical Informatics, University of Giessen

Birgit Samans
Faculty of Health Sciences, University of Applied Sciences

Keywan Sohrabi
Faculty of Health Sciences, University of Applied Sciences

Research Article

Keywords: Clinical Trial Recruitment Support System, Patient Selection, Logical Observation Identifiers Names and Codes, Interoperability, OMOP

Posted Date: June 1st, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2985787/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Background

Clinical trials are an essential part of the medical research process, but trials are often canceled due to lack of enrollment. Clinical trial recruitment support systems can help increase the number of participants by finding more suitable subjects. The ATLAS software (developed by OHDSI) can support the initiation of a clinical trial by building cohorts of patients who meet certain eligibility criteria. The correct use of medical classification systems for clearly defined inclusion and exclusion criteria in studies is an important pillar of this software.

Objectives

The purpose of this study is to determine whether ATLAS can be used in a Clinical Trial Recruitment Support System to represent the eligibility criteria of clinical trials. Therefore, we have considered the set of criteria implemented in ATLAS and realistically identified the strengths and problems of the software.

Methods

We analyzed ATLAS using 223 eligibility criteria from 30 randomly selected studies from the past 10 years. Therefore, appropriate ICD, OPS, LOINC, or ATC codes were selected during the data collection process to feed the software. Each criterion and study was ranked for feasibility of implementation in the software.

Results

According to the observations, 49.33% of the analyzed inclusion criteria are fully implemented in ATLAS. In our selected sample set, 10% of the studies are classified as fully portrayable and 73.33% as portrayable. In addition, the software is evaluated with respect to its usability and its interaction with medical classification systems.

Conclusions

To improve and extend the coverage of criteria in cohort definition in a real-world setting, we recommend working closely with the individuals involved in the study to precisely define the criteria and carefully select terminology systems. The selected criteria will be combined according to the individual setting. Further work is needed to specify the relevance and quantity of the extracted criteria.

1 Background

In the field of medicine, the goal of clinical trials is always to learn more about the human body in an effective and positive way. The ultimate aim of a medical study is always to improve patient care and therapeutic standards through the application of clinical trials. In order to observe the influence of the
new interventions and preventive measures in terms of safety and efficiency, a certain number of participants are recruited according to certain criteria. Such recruitment is always a challenging and time-consuming stage in the medical trial pipeline, resulting in a large number of canceled or incomplete trials (1–3).

In each clinical trial, eligibility criteria (eligibility criteria) are defined at the beginning of study development describing the relevant characteristics shared by the participants. Patient cohort identification and recruitment usually is carried out by research staff or primary care personnel through querying the clinical systems manually for patients matching the eligibility criteria. The latter process is a time-consuming and cost-intensive part of a clinical research study pipeline. Therefore, utilizing digital tools can significantly improve the recruitment of subjects and reduce costs and required labor resources (4).

Clinical Trial Recruitment Support Systems (CTRSS) or patient recruitment systems (PRS) can booster patient inclusion of clinical trials by automatically analyzing eligibility criteria based on electronic health records (4–6). Although these systems nowadays are integrated increasingly in many information and communication technology information and communication technology medical research projects, most of the study centers do not tend to use such digital tools for patient recruitment (7). The main reason is the requirement to domain technical staff as well as equipment to perform implementation or supportive tasks for the CTRSS. To address this issue, an individual system can be launched for multiple platforms or to integrate the CTRSS with the existing research systems.

In addition to the existing CTRSS or commercial solutions developed as part of different projects, several open-source tools like the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) developed by Observational Health Data Sciences and Informatics (OHDSI) or Informatics for Integrating Biology and the Beside (i2b2) established by National Institutes of Health, are free to use for patient recruitment. Both systems are based on the idea of using standard terminologies within data repositories however, they are not bound to a specific nomenclature. In addition, data can be extracted from different sources and in various models, so they may be utilized in different medical contexts and at multiple study centers with different prerequisites (8, 9). Therefore an architecture for a CTRSS utilizing OMOP as its data basis can be used for a patient recruitment system for example.(10, 11)

1.1 Observational Health Data Sciences and Informatics (OHDSI)

The OMOP CDM is part of a whole tool suite created and maintained by OHDSI, as a collaboration among 150 organizations around the world that collect and process healthcare data. OHDSI offers a wide range of open-source tools to support various data-analytics use cases on observational patient-level data, all are interacting with databases using the OMOP CDM. (8, 12)
Inspired by OHDSI, the ATLAS software was designed and developed as a web-based open source application. The primary goal of ATLAS is to be used for the observational analysis and generating real-world evidence through patient data extracted from clinical practice. ATLAS operates as a user interface on the OMOP CDM. According to (10) ATLAS allows defining cohorts, selecting analysis configurations, and tagging diseases with appropriate codes. Also, it allows sharing the project easily with other researchers (8).

ATLAS can also be used for subject recruitment by formalizing the eligibility criteria of a trial for the corresponding cohort. The definition of trials in ATLAS works with two main designs: cohorts and concepts. Concepts are individual codes that belong to a terminology system integrated in the OMOP CDM. They can be grouped into concept sets. A concept set contains several concepts from the standardized vocabulary in combination with logical indicators. It allows the user to specify whether related concepts should be included or excluded from the vocabulary hierarchy. Cohorts are the collection of individuals searched for by a particular query based on specific criteria, so all of a study’s eligibility criteria are part of the cohort definition (8).

### 1.2 Terminology systems

The database of ATLAS enforces the feeding data to follow standardized classification systems like Logical Observation Identifiers Names and Codes (LOINC) or Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT). These nomenclatures work as an interface for the study centers and built the basis for generating patient cohorts. (13, 14)

Several analyses have been conducted to investigate the use of digital systems based on international terminologies. (15, 16) For example, (17) investigated the use of a Clinical Data Interchange Standard Consortium (CDISC) based model in conjunction with a patient-centered model to structure eligibility criteria of clinical studies. International classification systems, such as International Statistical Classification of Diseases and Related Health Problems (ICD), formed the basis for this project. The authors used 200 randomly selected inclusion criteria from other studies, and the majority of the criteria could be represented in the system. (17)

Automatic tools like criteria2query use natural language processing (NLP) to automate formalization of eligibility criteria into the ATLAS cohorts. The latter system works just on specific terminology systems and local specialties are not considered. As the definition of an ATLAS cohort depends highly on the engaged terminology systems, such tools cannot be applied in the presence of other supported coding systems. (18–21)

Although there are more studies available addressing the latter issue, there still is a gap left between the theory and practically introducing the eligibility criteria with different terminology systems into the software (15, 22). Therefore, further analysis should be conducted regarding to this matter in order to explore the hidden potentials behind the different software and coding systems. This study considers the
possibility to apply inclusion and exclusion criteria of different clinical trials in ATLAS as a proof of concept.

2 Methods

2.1 Objectives

The first step in applying the eligibility criteria into ATLAS is to discover the feasibility of their implementation (in terms of quantity) with respect to the utilized standardized open-source terminology system. Also, the technical difficulties in the implementation process will be analyzed as the qualitative drawbacks.

The selected terminology systems in our study are the standard classification systems ICD and Operation and Procedure Code (OPS), which is a German modification of the International Classification of Procedures in Medicine (ICPM), utilized for billing in all German hospitals. Also, the Anatomical Therapeutic Chemical classification (ATC) system is selected for the medications and LOINC coding system for laboratory results as well.

In a next step, all eligibility criteria which could not be introduced into ATLAS are considered. They are categorized based on the shortlisted drawbacks above.

This paper works on establishing classification systems and standards, as well as dealing with the mapping of inclusion and exclusion criteria generally and in ATLAS especially.

2.2 Study design

To initiate the research, we randomly selected 30 studies without any preference with respect to a medical discipline. For this sake, we searched on the platform "ClinicalTrials.gov", one of the largest registries of clinical trials including about 425,817 studies from 222 different countries (23). The selected studies had all to be started within the last 10 years resulted in 114 studies wherein 30 studies have been selected randomly. Also, to investigate the extent to which the eligibility criteria can be implemented in ATLAS, we documented all 331 eligibility criteria of the 30 studies in a table.

2.3 Procedure in ATLAS

To begin the process of introducing criteria into ATLAS, we selected an appropriate terminology system with appropriate codes for each eligibility criteria, as shown in Fig. 1. The appropriate codes were identified using official online catalogs (13, 24–26).

If no appropriate codes were found in these catalogs, SNOMED CT would have added to the analysis to check for an alternative way to introduce or embed an approval criterion. (14)

In ATLAS all fitting codes describing a single eligibility criterion are summarized in one concept set, which is a collection of codes e.g. concepts. This is necessary, because a single criterion can be constituted of
one up to hundreds of codes. For example, type I diabetes mellitus can have different characteristics, which can be marked with different ICD codes. Codes for these disease individualities create one concept set. In comparison, Hemophilia A could be created as a concept set using a single code, as there is only one ICD code for this specific disease (Fig. 2). If there was any doubt about an eligibility criterion or related codes, medical staff was consulted. Afterwards, a separate concept set had to be created for each eligibility criterion, although no concept sets were created for demographic data such as age or gender. The aforementioned criteria were entered directly during the later process within a cohort by means of the corresponding functions available in ATLAS.

Once all the concept sets required for a study were established, cohorts, that included all eligibility criteria for a study, had to be defined. To outline a cohort, at first it is necessary to define an initial event, which is a main criterion and used for a preselection of subjects. All individuals, which fulfill the requirements of the initial event, build the main cohort of the patients. To set further constraints, it is possible to set other criteria to get more specified results. We always selected the examined disease of the trial as an initial event for the cohort. For example, if a study needed test group who meet selection criteria such as type I diabetes mellitus, age under 18 years and elevated Body-Mass-Index (BMI), we would define diabetes mellitus as initial event. All patients with type I diabetes built the main cohort. Criteria “age” and “BMI” would have added as further eligibility criteria.

3 Evaluation of the studies and eligibility criteria

3.1 Categorization of eligibility criteria and studies

Of all 331 eligibility criteria, 108 criteria were excluded because of missing relevance. These criteria were the ones, which do not exist prior to a face-to-face interview with a study physician so, they cannot be analyzed in the process of the electronic patient screening.

To categorize the eligibility criteria, we analyzed the remaining 223 suitability criteria and classified them in terms of their implementation in ATLAS into the categories implementable, partially implementable, and not implementable (s. Table 1). If a criterion was only partially or not implementable, we would also add short notes expressing our decision.
Table 1
Criteria categories

<table>
<thead>
<tr>
<th>Criteria Category</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>implementable</td>
<td>Criterion can be implemented in ATLAS with given terminology systems.</td>
</tr>
<tr>
<td>partially</td>
<td>Criterion can be implemented, but not with the given terminologies (only SNOMED CT) or only a part of criterion is mapped (for example a disease can be implemented, but a timely constraint is not possible).</td>
</tr>
<tr>
<td>not implementable</td>
<td>Criterion cannot be implemented with the help of the given terminologies or SNOMED CT and the functions available in ATLAS.</td>
</tr>
</tbody>
</table>

After weighting the studies with respect to their feasibility, four categories have been established: fully portrayable, portrayable, partially portrayable, and not portrayable. (see Table 2).

Table 2
Trial Categories

<table>
<thead>
<tr>
<th>Trial Category</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>fully portrayable</td>
<td>All criteria of a study can be implemented in ATLAS.</td>
</tr>
<tr>
<td>portrayable</td>
<td>The initial event is classified as implementable and can be mapped completely in ATLAS. Further inclusion criteria can be classified as partially implemented or not implementable according to Table 1.</td>
</tr>
<tr>
<td>partially portrayable</td>
<td>The initial event is classified as partially implementable in ATLAS according to Table 1.</td>
</tr>
<tr>
<td>not portrayable</td>
<td>The initial event cannot be implemented in ATLAS.</td>
</tr>
</tbody>
</table>

Explanation of trial categories in the overall evaluation about feasibility of measures taken in ATLAS

3.2 Classification of partially/not implementable criteria

Finally, we made a secondary analysis to discover the main difficulties in the introducing into or implementation of eligibility criteria in the software ATLAS. Therefore, all reasons and argumentations have been considered to justify the categorization and the implementation issues of the eligible criteria.

3.3 Ethical considerations

No ethical vote was needed.

4 Results

The described categorization resulted in an implementation of 30 cohorts with one initial event per cohort and an additional number of 223 eligibility criteria. For implementing all eligibility criteria and initial
events, we built 129 different concept sets, including 11.2 concepts in average. Each cohort included in average 5.33 concept sets with 59.73 concepts in total. The whole table can be found in Appendix 1.

4.1 Evaluation of eligibility criteria

According to our observation, 49.33% (n = 110) of all eligibility criteria identified in the studies could be mapped in ATLAS so classified as implementable (see. Table 1). This set is corresponding with the eligibility criteria referring to clearly defined diseases, procedures, or drugs respectively with the corresponding ICD, OPS, ATC or LOINC codes. Also, 20.63% (n = 46) of the criteria were defined as partially implementable and 30.04% (n = 67) could not be implemented in ATLAS at all.

4.2 Overall evaluation of analyzed studies

Based on our evaluation, in 10% of the studies (n = 3) all criteria could be fully mapped, so these studies were classified as fully portrayable, (see. Table 2). The latter set includes a few simple and clearly defined criteria, such as a well-defined diseases, forming a group of subjects which could be filtered more by one or two additional criteria, like age or gender. In these cases, the disease has been introduced into ATLAS using the appropriate codes and as an initial event.

Also, 73.33% of the studies (n = 22) have been classified as portrayable, where at least one of the eligibility criteria was only partially or not implementable according to the rules in Table 1. However, the initial event could be fully implemented as a concept set using ICD, OPS, LOINC or ATC. In 16.67% (n = 5) it was not possible to fully implement the initial event and these were assigned to the category partially portrayable. Finally, none of the studies was classified as not portrayable.

4.3 Partially implementable criteria

After analyzing the justifications corresponding to the criteria which could not be introduced into ATLAS, we have categorized them in four groups described in Table 3.

<table>
<thead>
<tr>
<th>Criteria Category</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing information</td>
<td>Description is not precise</td>
</tr>
<tr>
<td>Highly specialized criterion</td>
<td>Highly specialized diagnosis, procedure or laboratory result</td>
</tr>
<tr>
<td>Technical reasons</td>
<td>Partial aspect is not feasible due to technical reasons</td>
</tr>
<tr>
<td>Terminology system</td>
<td>Only implementable with SNOMED CT</td>
</tr>
</tbody>
</table>

According to the results, 39.13% (n = 18) of the partially implemented criteria are belonging to the associated class due to technical reasons, with different causes. For example, in some studies, the ages are quoted in years but in others like pediatric studies dealing with the infants quoted in a monthly basis. In ATLAS it is only possible to query the age in years. Some criteria also need a cause-and-effect
relationship. In such criteria, the information whether the patient is involved in cross studies is required which cannot be extracted from an OMOP data model.

Among the rest of the criteria, 26.09% (n = 12) could not be fully implemented due to missing information. Those criteria mainly refer to “norm” or “stable” values even if there are no defined norm quantities valid in all countries or locations but different standards from site to site. The latter set includes criteria that refer to the groups of medicaments, diagnosis, procedures, or laboratory results, which are not globally defined.

Another set of criteria includes 17.39% (n = 8) criteria which were too specific and no terminology system has been found to have a specific corresponding code, except for the codes that also include a group of diseases/procedures/laboratory/results/medications. Finally, 17.39% (n = 8) eligible criteria, have been classified with SNOMED CT.

4.4 Not implementable criteria

To categorize the criteria which are not implementable, we have selected the same categories, without the category corresponding to the wrong terminology system. In 47.76% (n = 32) eligible criteria, missing information leads to the categorizing as “not implementable”. 28.36% (n = 19) of the criteria were too specific and 23.88% (n = 16) were not implementable due to technical reasons.

4.5 Technical usability aspects of ATLAS

According to the practices performed with ATLAS in this work, it is observed that the intuitive operations in the software have a poor performance which necessitates a training phase prior launching the main study work. Individual adaptations are needed to enter complex or nested study criteria. For example, the scores cannot be noted by selecting an available function in ATLAS or by introducing one specific code as a concept set. The analyzed scores usually consist of several sub-items, such as diseases or laboratory values, which are assigned a value depending on their severity. In these cases, the content of each score first has to be reviewed more in detail. This means that all aspects of a score are treated in the first step like individual criteria that have to be introduced into the system as separate concept sets, which is a major effort for complex scale studies.

ATLAS has the possibility to input basic demographic data in an easy way and without assigning a concept, but these data are limited to age and gender. Other basic information like height and weight has to be input like other observations. For this purpose, the appropriate LOINC code has to be identified, created as a concept set and time constraints might have to be set, when only the latest data on height and weight are inquired. At this point, this information can be integrated into the cohort.

Also, temporal information, like the starting point of an event or a treatment can only be introduced as an absolute date, not as relative time information like “six month ago”. Therefore, the selection criteria for a cohort of a time period must be updated on a daily basis.
5 Discussion

5.1 Digital subject recruitment

OMOP and ATLAS as the products of OHDSI have been developed to support multi-institutional research projects. In this study, we intended to test the possibility of using OMOP and ATLAS not only to run analysis on the existing data sets, but also to portray the eligibility criteria of clinical trials. This makes it suitable as a digital tool for recruiting of candidates for clinical studies. We have observed that the implementation of trial eligibility criteria in ATLAS cohorts is more complex than a 1:1 translation: about half of the eligibility criteria could not be fully implemented in ATLAS.

One main reason (n = 44) is that the eligibility criteria from study descriptions are defined for the use of study physicians or study personnel and often leave scope for individual professional assessment. These are easy to interpret for medical domain experts while reading the criterion, such as “stable laboratory results”, in an intuitive way. However, these criteria are not globally defined and each physician can have a slightly different interpretation of “stable” for example. A deep understanding of the criteria is essential for the process and the lack of such an understanding, results in criteria not being introduced into the system or being incorrectly implemented. The precise formulation of subject characteristics is required for a clinical research initially, so well-chosen criteria are crucial for the feasibility and internal validity of studies (27). In terms of ATLAS, carefully chosen and precisely formulated eligibility criteria form the basis for being able to use the software correctly and to exploit its potential functionalities. To get better results in the process of formalizing eligibility criteria for an automated search, it is helpful to collaborate tightly with the study personnel, so that imprecise criteria can be refined by professionals. Subsequently, all these criteria and therefore more trials, can be implemented in ATLAS.

5.2 The utilized terminology systems

Since the ATLAS software is based on the involvement of terminology systems, the analysis highlights the importance of the choice of terminology systems and thus the availability of codes contained in terminology system. ICD, ATC and LOINC are international classification systems widely used in German clinics in hospital information systems and therefore they are easy to access and use if hospital data are queried. Also, OPS is a terminology system only used in Germany applied for the code procedures for billing (28). In addition to the 49.33% of the terms that could be mapped by ICD, ATC, LOINC and OPS, further 4% of the criteria could be mapped by using SNOMED CT, which is the standard terminology system used in OMOP (8).

Terminology systems have a huge impact on the opportunities in building concept sets and cohorts in ATLAS. The medical terminology in SNOMED CT is significantly in more details than in ICD. SNOMED CT includes not only diagnoses itself, but also medical terms, which are placed in relation to each other. The use of ICD is recommended to record the diagnoses for billing and has a hierarchically arranged structure of diagnoses, so it is more widely used in the patient care systems and therefore provides the ease of access. Consequently, it is essential to inspect the utility and target group before making a statement.
about nomenclatures being better suited for a specific purpose. A combination of the coding methods, for example combination of ICD and SNOMED CT, is also recommended by (29).

However, there were also many exclusion and inclusion criteria, which could not be mapped by means of chosen classification systems (neither by ICD, OPS, ATC, or LOINC, nor by SNOMED CT) due to high individuality of the subjects (n = 27). Clinical trials are often designed to gain knowledge about novel or under-researched diseases. The codes for these conditions do not yet exist in the medical classification systems, which makes it difficult or impossible to introduce some selection criteria into ATLAS. In some cases, eligibility criteria refer to non-medical settings like the proof of a vaccination. Those criteria are not commonly included in the medical terminologies, as the corresponding domain is a medical and not an administrative scenery.

According to our observations, we can conclude that the formulated eligibility criteria have an impact on using correct terms, but on the other hand, for the specialized criteria no fitting codes are available. Also, it has been shown that the choice and correct use of medical classification systems, like the definition of inclusion and exclusion criteria, is an important basis for working with ATLAS. Therefore, there is a need for the correct use of nomenclatures and standardization, as well as continual adaptations to medical progress, which researchers, physicians, and others involved in health care need to strive for.

In this study, the effectiveness and suitability of the individual classification systems has not been analyzed. So, we cannot state a verdict on the usability of some nomenclatures in ATLAS.

5.3 Limitations of ATLAS

As expressed before, the requirement to a training phase before working with the software is a drawback. Also, in some cases, individual adaptations are required to introduce complex or nested data sets such as scores. This indicates problems with usability of ATLAS that have been also confirmed by other studies. In 2021 (30) focused on the timeline availability of clinical data and bio specimens that can be identified using digital applications. Three systems, i2b2, OHDSI's ATLAS, and Sample Locator by the German Biobank Alliance, were tested in terms of strengths, weaknesses, and associated usability via qualitative exploratory study followed by a web-based usability test. The results of this study show that on one hand ATLAS has significantly more functionalities than other tools and a very good interface visualization. However, multiple usability issues exist as well. These problems can be associated to the large number of functions and different approaches when introducing selection criteria. However, the authors take into account that ATLAS has been primarily developed for the researchers and experts who would often define complex cohort queries. Therefore it is important to provide them with the variety of selection and input options, however, it increases the operational costs (30).

This study also suggests that ATLAS is more suitable for people with the appropriate technical background, due to the existing features and the need for a certain level of development for the sake of customization. However, according to the developers, the application aims to provide the greatest scientific impact for researchers. Therefore, promoting the usability and adaptability of the software for
this target group is of great importance. This should be supported by members of the target group as described in (31) wherein the medical students and licensed physicians were asked to test electronic health record used in Finland in terms of usability. It has been concluded that the physicians were able to identify potential problems corresponding to the usability aspect and provide the corresponding feedback. So, it is recommended that target groups of a specific application to be involved in the usability development process in the test phases to detect problems in a timely manner and avoid stress in the future.

It is worth mentioning that an inaccurate implementation of eligibility criteria in ATLAS can lead to less precise subject recommendations. If the entry of studies into ATLAS is intended to lead to suitable subject recruitment, the involvement of specialized study physicians or study personnel in the process of cohort definition will be helpful.

6 Conclusions

In this study, we have considered the selection criteria of 30 studies for the possibility to portray them in the software ATLAS. We have found out that about 50% of the criteria could be implemented into the software, about 20% could only be partially implemented, and the rest of 30% could not be implemented at all.

The four main justifications are the existence of missing information about the criteria, limitations of the software, limitations of the chosen terminology systems and finally highly specialized criteria. To use ATLAS as a CTRSS, we recommend the following points to improve the coverage of eligibility criteria in a cohort definition.

As ATLAS offers a variety of holistic functionalities, the usability degree is not trivial. It can be useful to plan a training phase before starting with the main work. Even for the planning phase of a project, it is helpful to know all possibilities and limitations of ATLAS right from the start. Another difficulty is the limitation of the selected terminology systems. It can be helpful to examine the options and possibilities of different terminology systems to find the best and suitable system for the project. The combination of different terminology systems can also be useful to improve the coverage of eligibility criteria in cohort definitions.

The main problem in portraying clinical trials in ATLAS is the missing information in eligibility criteria. A tight cooperation with study personnel, which is usually responsible for the subject recruitment and works with the criteria daily, can help to specify these uncertainties. If all missing information can be cleared with the study personnel, this could raise the percentage of implementable criteria up to about 70%. This could help with criteria that are highly specialized to be covered by terminology systems, and therefore it is only possible to implement parts of the criteria or use collective codes, which are not precise. Usually, the study personnel has similar issues when searching for eligible persons, so they can help in identifying other ways to portray these criteria. Even if the criteria are not fully implementable, it is sometimes possible to use similar criteria, which are valid for a preselection. In addition, the study personnel can
prioritize the relevance of different criteria, as some of the criteria are more important for a selection of candidates than others.

In this work, we picked one of the criteria as the main criterion for the cohort, and leave the rest with no prioritizing. Although some criteria are more relevant for the selection process of the subjects than others. The importance level of a criterion, can only be estimated by the study personnel working in the recruitment process. Further work is planned to examine which and how many of the criteria are relevant for the selection process of clinical trials and to survey how many of the relevant criteria can be implemented.

Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATC</td>
<td>Anatomical Therapeutic Chemical classification</td>
</tr>
<tr>
<td>BMI</td>
<td>Body-Mass-Index</td>
</tr>
<tr>
<td>CDISC</td>
<td>Clinical Data Interchange Standard Consortium</td>
</tr>
<tr>
<td>CDM</td>
<td>Common Data Model</td>
</tr>
<tr>
<td>CTRSS</td>
<td>Clinical Trial Recruitment Support System</td>
</tr>
<tr>
<td>ICD</td>
<td>International Statistical Classification of Diseases and Related Health Problems</td>
</tr>
<tr>
<td>ICPM</td>
<td>International Classification of Procedures in Medicine</td>
</tr>
<tr>
<td>i2b2</td>
<td>Informatics for Integrating Biology and the Beside</td>
</tr>
<tr>
<td>LOINC</td>
<td>Logical Observation Identifiers Names and Codes</td>
</tr>
<tr>
<td>OHDSI</td>
<td>Observational Health Data Sciences and Informatics</td>
</tr>
<tr>
<td>OMOP</td>
<td>Observational Medical Outcomes Partnership</td>
</tr>
<tr>
<td>OPS</td>
<td>Operation and Procedure Code</td>
</tr>
<tr>
<td>PRS</td>
<td>Patient Recruitment System</td>
</tr>
<tr>
<td>SNOMED CT</td>
<td>Systematized Nomenclature of Medicine Clinical Terms</td>
</tr>
</tbody>
</table>

Declarations

Ethics approval and consent to participate
Not applicable

Consent for publications
Not applicable
Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files.

Competing interests

The authors declare no conflict of interest.

Funding

This work is part of the project MIRACUM, funded by the German Ministry of Education and Research (FKZ 01ZZ1801D).

Authors’ contributions

RB and KMB wrote the original texts of this work. RB and KS did the conceptual work of this study. KMB analyzed all trials and criteria, as well as implemented and categorized the criteria. HS and BS were involved in the conceptual work of this article. All authors read and approved the final manuscript.

Acknowledgements

We are thankful to Dr. rer. nat. Seyed Amir Hossein Tabatabaei for reviewing this work and offering useful advice.

References


Figures
**Figure 1**

Procedure in ATLAS
### Concept Set of:
**Type I diabetes mellitus**

- E10.0: Type 1 diabetes mellitus with coma
- E10.1: Type 1 diabetes mellitus with ketoacidosis
- E10.2: Type 1 diabetes mellitus with renal complications
- E10.3: Type 1 diabetes mellitus with ocular complications
- E10.4: Type 1 diabetes mellitus with neurological complications
- E10.5: Type 1 diabetes mellitus with peripheral vascular complications
- E10.6: Type 1 diabetes mellitus with other specified complications
- E10.7: Type 1 diabetes mellitus with multiple complications
- E10.8: Type 1 diabetes mellitus with unspecified complications
- E10.9: Type 1 diabetes mellitus without complications

### Concept Set of:
**Hemophilia A**

- D66: Hereditary factor VIII deficiency

---

**Figure 2**

Comparison of concept sets between type I diabetes mellitus type (left) and hemophilia A (right)

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- Appendix1AllTrialsandCriteriawithcategories.xlsx