German Medical Data Sciences: A Learning Healthcare System
U. Hübner et al. (Eds.)

doi:10.3233/978-1-61499-896-9-3

Investigating the Capabilities of FHIR Search for Clinical Trial Phenotyping

Christian GULDEN [a,1], Sebastian MATE [a], Hans-Ulrich PROKOSCH [a] and Stefan KRAUS [a]
[a] *Medical Informatics, Univ. of Erlangen-Nürnberg, Erlangen, Germany*

Abstract. Clinical trials are the foundation of evidence-based medicine and their computerized support has been a recurring theme in medical informatics. One challenging aspect is the representation of eligibility criteria in a machine-readable format to automate the identification of suitable participants. In this study, we investigate the capabilities for expressing trial eligibility criteria via the search functionality specified in HL7 FHIR, an emerging standard for exchanging healthcare information electronically which also defines a set of operations for searching for health record data. Using a randomly sampled subset of 303 eligibility criteria from ClinicalTrials.gov yielded a 34 % success rate in representing them using the FHIR search semantics. While limitations are present, the FHIR search semantics are a viable tool for supporting preliminary trial eligibility assessment.

Keywords. FHIR, phenotyping, clinical trials

1. Introduction

Clinical trials are the foundation of evidence-based medicine; they are used to evaluate new treatment options such as drugs or surgical approaches in a well-defined and controlled environment. Whether a given patient can be considered a suitable participant for a trial depends on the extent by which his demographic and clinical characteristics match a set of eligibility criteria defined by the designers of the trial. In order to search for eligible patients for a study, a variety of clinical trial recruitment support systems (CTRSS) have been developed, which, based on implicitly or explicitly formalized in- and exclusion criteria, can query for potential trial patients from data available in the electronic health record [1].

The Health Level 7 Fast Healthcare Interoperability Resources (HL7 FHIR) standard is being developed to address the challenges of an increasingly digitized healthcare ecosystem. The specification leverages emerging and well-established industry standards for exchanging data between healthcare applications based on the experiences gained and lessons learned from developing previous standards such as HL7 v2, v3, and Clinical Document Architecture (CDA) [2]. FHIR defines a set of resources representing common clinical objects and concepts as well as operations for creating, modifying, and searching for them based on defined attributes. In FHIR, resources are considered the basic building blocks of integration. They generally represent clinical concepts such as patient, medication, and practitioner. Every resource has a set of attributes: a *patient*

[1] Corresponding Author, Christian Gulden, Medical Informatics, Friedrich-Alexander Universität Erlangen-Nürnberg (FAU), Wetterkreuz 13, 91058 Erlangen, Germany; E-mail: Christian.Gulden@fau.de.

resource, for example, encompasses demographic data such as the date of birth and gender of a patient and an *observation* resource contains information about a lab test or measurements compiled by a clinical device. Such resources may reference – or may be referenced by – other resources.

The RESTful create, read, update, and delete (CRUD) operations of the FHIR API are complemented by a search functionality which allows for filtering the set of resources by parameters supplied to the HTTP URL query [3]. In this study, we investigate the capabilities of these search semantics and demonstrate how this constraint-based search can be used to formulate clinical trial inclusion and exclusion criteria to find eligible patients. We limit the scope of this work by exclusively focusing on the features provided by a standard-conformant FHIR server and resources out-of-the-box and not on the role FHIR may play as part of a larger CTRSS.

2. Methods

As input for our investigation and experimental validation, we downloaded the HTML representation of the 100 latest (as of January 1st 2018) trials that were open for recruitment from ClinicalTrials.gov and randomly selected a subset of 25. We created a tool to automatically parse and extract the eligibility criteria from this data. The tool considers each bullet point in the 'Eligibility Criteria' section of the HTML page describing the trial as a single criterion.

We investigated in how far this subset of criteria can be encoded using the search functionality provided by the FHIR DSTU3 specification as implemented by a HAPI-FHIR server[2]. In order to retrieve the final set of patients from the inclusion and exclusion criteria, we created an additional tool which receives the FHIR query representations of both sets of criteria as input and returns a final set of eligible patients by excepting the set of patients fulfilling the exclusion criteria from those fulfilling the inclusion ones.

3. Results

From the 25 trials, we extracted a total of 303 eligibility criteria; 119 describe inclusion conditions, 184 exclusions. The mean number of inclusion criteria per study was 4.76 (IQR=4, Range=1-12) and for exclusion criteria it was 7.36 (IQR=9, Range=1-20).

Table 1. Numerical overview of the analyzed criteria and their representability.

	Representable	Not Representable	Total
Inclusion Criteria:	48	71	119
Exclusion Criteria:	56	128	184
Total:	104 (34 %)	199 (66 %)	303

Table 1 shows the statistical results of the encoding process. A criterion was considered representable if it described a requirement that could be expressed by the syntax and semantics of the FHIR search. In the following, we will give some examples of these representations and the limitations we encountered.

[2] http://hapifhir.io/

Table 2. List of trial inclusion criteria and their query representation of one clinical trial[3].

No.	Criterion	URL Query Representation
1	Diagnosis of a major depressive disorder	/Condition?code=http://snomed.info/sct|370143000
2	Age greater or equal to 18 and less or equal to 34	/Patient?birthdate=le2010&birthdate=ge1984
3	Ability to fluently read, write, and speak Dutch	/Patient?language=nl
4	Sexes Eligible for Study: Female	/Patient?gender=female
5	Logical conjunction of all criteria	/Patient?gender=female &birthdate=le2010&birthdate=ge1984 &language=nl &has:Condition:patient:code =http://snomed.info/sct|370143000

The degree by which the search methodology defined in the FHIR standard can be used for representing clinical trial eligibility criteria strongly depends on the kind of data requested in the criteria, the kind of data available in the queried resource, as well as the semantic complexity of the criteria. The query API only supports simple numerical or categorical comparisons when filtering resources. Table 2 shows the inclusion criteria for one of the studies we analyzed and their representation as a search query. In order to now find all patients that match these inclusion criteria, we combine them using the logical AND for a composite search of patient resources resulting in the final query (5).

The example trial in Table 2 demonstrates the general category of criteria which can easily be represented using the FHIR RESTful API search syntax: categorical comparisons (1, 3, 4) and simple numerical range evaluations (2) using quantitative prefixes (`le` = less or equal, `ge` = greater or equal, etc.). Typical examples of these criteria relate to demographic characteristics, such as patient age, gender and basic diagnosis. They are commonly present in the trial criteria as well as the underlying electronic health record (EHR) data source [4].

4. Discussion

The FHIR RESTful search provides an easy way to search for patients based on common inclusion and exclusion criteria. Using FHIR for phenotyping leverages the capabilities of an open standard, thereby easing the reuse and exchange of these queries.

In Table 2 criterion (2) shows that by default the patient resource only exposes the birthdate attribute as a valid search parameter. This makes it necessary to explicitly define the age as a period relative to the current year. However, resources provide the ability to be extended by custom attributes and search parameters. For example, the Patient HL7 profile adds an extension with additional search parameters such as age to the patient resource [5], thus simplifying the criteria query to `/Patient?age=ge18&age=le34`. This allows for the query to be reused over time without modifications. This problem not only applies to the patient age, but also more generally to any situation in which absolute duration information is required from timestamps stored in FHIR resources. For example, exclusion criteria often require eligible candidates to not have undergone a specific form of therapy, or have taken some

[3] https://clinicaltrials.gov/show/NCT03388177

medication, within the past N days prior to the study begin. Extensions of resources can thus simplify constraint-based searches. However, which extensions are required to better support the search semantics for phenotyping in the context of clinical trials remains an open question.

When attempting to represent more complex eligibility criteria, one quickly runs into the limitations of the FHIR constraint-based search approach. For example, one study[4] aims to include children whose BMI is less than the 95th percentile for their age, with children between 2 and 5 years old considered for the study. In this case, one would first have to determine the age of a child, potentially calculating the BMI based on weight and height, or retrieving it from extension of observation resources, and finally determining whether the BMI is less than the 95th percentile for that particular age. Such queries, with inter-data dependencies and necessary computations, cannot be expressed.

Table 3. Example of a complex temporal inclusion criterion and the required explicit exclusion criterion necessary for correctly identifying eligible patient cohorts.

No.	Criterion	Query Representation
1	Patients who are 1 year post liver transplant, but within 5 years of transplant	/Patient?_has:Procedure:patient:code =http://snomed.info/sct|274025005 &_has:Procedure:patient:date=ge2013 &_has:Procedure:patient:date=le2017
2	Explicit exclusion criteria	/Patient?_has:Procedure:patient:code =http://snomed.info/sct|274025005 &_has:Procedure:patient:date=ge2017

Furthermore, complex temporal constraints are in general difficult to represent. Instead, only basic time ranges can be expressed. One complex inclusion criterion is shown in Table 2. Here, we are searching for patients who had a liver transplant performed on them with the temporal constraint of it happening at least one year ago but not more than five. While the query may seem correct at first, this approach shows one limitation of the constraint-based search: a patient who had a liver transplant in 2015 and an additional transplant mid-2017 would be included by criteria (1) even though the patient is not fully eligible. In order to avoid such a scenario, an additional explicit exclusion criterion (2) has to be defined, executed and the intersecting elements discarded.

On the topic of exclusion, the FHIR search standard only provides semantics for declaring that resources must *not* fulfill a given constraint for search parameters of type number, date, and quantity (e.g. using a 'not equals'-prefix) and not, for example, for disease coding. This makes it difficult to generally represent both inclusion and exclusion criteria in a single query. Instead, intersecting patients between the inclusion and exclusion criteria need to be explicitly removed from the final result set. To address this, we used the second tool described above to build the set of patient resources by excepting the set of excluded patients from the set of included ones.

The HL7 Clinical Quality Language (CQL) [6] offers a much richer set of features for formalizing, among others, trial eligibility criteria. This richness requires stronger familiarity with its constructs as opposed to the RESTful paradigm which is in wide use even beyond the healthcare sector. Included in CQL and specified in FHIR is also FHIRPath, a path-based navigation and extraction language. It defines mathematical

[4] https://clinicaltrials.gov/show/NCT03394326

operations, subsetting, filtering, and additional functions for data processing. Evaluating the potential of FHIRPath for clinical trial phenotyping is subject for future work.

5. Conclusion

In this paper, we have investigated how the standardized FHIR RESTful search operations and capabilities can be used for clinical trial phenotyping by representing eligibility criteria as a set of constraint-based filters on FHIR resources.

While the FHIR standard does not explicitly provide the ability to query patients based on trial eligibility criteria, the available operations are helpful in supporting prescreening efforts in order to reduce the set of candidates based on common, expressible criteria prior to manual verification by medical professionals.

6. Conflict of Interest

The authors declare that there are no conflicts of interest in this work.

7. Acknowledgments

This research has been conducted within the MIRACUM consortium. MIRACUM is funded by the German Ministry for Education and Research (BMBF) under the Funding Number FKZ 01ZZ1801A. The present work was performed in (partial) fulfillment of the requirements for obtaining the degree “Dr. rer. biol. hum.” from the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) (CG).

References

[1] M. Cuggia, P. Besana, D. Glasspool, Comparing Semi-Automatic Systems for Recruitment of Patients to Clinical Trials, *International Journal of Medical Informatics* **80** (2011), 371-88.
[2] HL7 FHIR, Available at: https://www.hl7.org/fhir/, (accessed 25 Jan 2018).
[3] HL7 FHIR, Search - FHIR v3.0.1, https://www.hl7.org/fhir/search.html, (accessed 23 Jan 2018).
[4] B. Trinczek, F. Köpcke, T. Leusch, R.W. Majeed, B. Schreiweis, J. Wenk, B. Bergh, C. Ohmann, R. Röhrig, H.U. Prokosch, M. Dugas, Design and Multicentric Implementation of a Generic Software Architecture for Patient Recruitment Systems Re-using Existing HIS Tools and Routine Patient Data, *Applied clinical informatics* **5** (2014) 264-83.
[5] HL7 FHIR, Patient-extensions - FHIR v3.0.1 https://www.hl7.org/fhir/patient-extensions.html, (accessed 23 Jan 2018).
[6] HL7, Clinical Quality Language, http://www.hl7.org/implement/standards/product_brief.cfm?product_id=400, (accessed 23 Apr 2018).