Contemporary Clinical Trials Communications 21 (2021) 100692

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

journal homepage: http://www.elsevier.com/locate/conctc

Assessing an Electronic Health Record research platform for identification of clinical trial participants

Niina Laaksonen [a,*], Juha-Matti Varjonen [b], Minna Blomster [b], Antti Palomäki [c], Tuija Vasankari [c], Juhani Airaksinen [c], Risto Huupponen [a], Mika Scheinin [a], Juuso Blomster [a,c]

[a] Institute of Biomedicine, University of Turku, Kiinamyllynkatu 10, FI-20520, Turku, Finland
[b] Auria Clinical Informatics, Hospital District of Southwest Finland, PO Box 52, FI-20521, Turku, Finland
[c] Heart Centre, Turku University Hospital, PO Box 52, FI-2052, Turku, Finland

ARTICLE INFO

Keywords:
Patient identification
Patient recruitment
Clinical trials
Electronic health records

ABSTRACT

Electronic health records (EHR) are a potential resource for identification of clinical trial participants. We evaluated how accurately a commercially available EHR Research Platform, InSite, is able to identify potential trial participants from the EHR system of a large tertiary care hospital. Patient counts were compared with results obtained in a conventional manual search performed for a reference study that investigated the associations of atrial fibrillation (AF) and cerebrovascular incidents. The Clinical Data Warehouse (CDW) of Turku University Hospital was used to verify the capabilities of the EHR Research Platform.

The EHR query resulted in a larger patient count than the manual query (EHR Research Platform 5859 patients, manual selection 2166 patients). This was due to the different search logic and some exclusion criteria that were not addressable in structured digital format. The EHR Research Platform (5859 patients) and the CDW search (5840 patients) employed the same search logic. The temporal relationship between the two diagnoses could be identified when they were available in structured format and the time difference was longer than a single hospital visit.

Searching for patients with the EHR Research Platform can help to identify potential trial participants from a hospital's EHR system by limiting the number of records to be manually reviewed. EHR query tools can best be utilized in trials where the selection criteria are expressed in structured digital format.

1. Introduction

Randomized clinical trials constitute a cornerstone for evidence-based medicine. In all clinical trials, the identification and recruitment of participants are critical steps for successful trial conduct. Poor patient recruitment is one of the most essential causes of delays in prospective clinical trials [1,2]. Only one in three trials reaches its recruitment target in the pre-defined timeframe, whereas the remaining two thirds of all trials either never reach the targeted sample size or the trial duration is extended because of slower than anticipated recruitment [2–5]. If the recruitment could be optimized by better selection of clinical study sites that have the potential and an identifiable patient population, and if more feasible clinical trial protocols could be created through improved and more accurate study site analysis, many clinical trials could be completed faster and more cost-efficiently.

To enhance the identification and recruitment of participants into clinical trials, new tools have been created for convenient, data-driven insights in the available patient population. Due to recent advances in Electronic Health Record (EHR) query technology, the use of EHR data for patient recruitment is currently seen as a promising tool to improve identification and recruitment of trial participants [6,7]. An EHR is an individual patient record contained within a hospital's electronic patient record system. A typical individual EHR may include the patient's medical history, diagnoses, treatment plans, immunization dates, medication records, and laboratory and other test results, including those derived from imaging investigations [8]. Some EHR data are presented in a structured format, whereas other data exist in the unstructured data fields of the EHR system. One example of an automated EHR Research tool created in Europe is InSite, a research platform maintained by Custodix N·V., Belgium. Through the platform, a hospital's EHR data can be utilized for research purposes, for example, for validating clinical trial protocols and for identifying patients who are

* Corresponding author. Institute of Biomedicine, University of Turku, Kiinamyllynkatu 10, FI-20520, Turku, Finland.
E-mail address: niinie@utu.fi (N. Laaksonen).

https://doi.org/10.1016/j.conctc.2020.100692
Received 6 May 2020; Received in revised form 5 November 2020; Accepted 14 December 2020
Available online 18 December 2020

potentially eligible for clinical trials [9]. Thus, the users can form queries based upon the trial protocol's inclusion and exclusion criteria and obtain counts of patients who match these criteria.

Traditionally, detection and preselection of eligible patients for clinical trials has been done by identification with manual searches from electronic patient records. This 'manual search' is able to provide reliable and controlled results and is considered a standard method for patient identification, but is typically a time and resource consuming method. In earlier evaluations, the InSite platform has been shown to be technically applicable across hospitals [9], but information on its performance characteristics compared with conventional manual search methods from a hospital's EHR system is limited.

2. Objectives

The aim of the present study was to evaluate how accurately an EHR Research Platform (InSite) can identify the same patients as discovered by a manual search in a reference study, and to evaluate the reasons for possible discrepancies. Secondly, the aim was to examine the capability of the methods to find the dates when atrial fibrillation (AF) and the corresponding index events (stroke, transient ischemic attack (TIA) or bleeding) were diagnosed and to identify the temporal relationship of the two diagnoses (i.e. which one occurred first, AF or the index event). To compare patient-level data and to assess the functionality of InSite as a digital search tool, the information and search engines of the Turku University Hospital's Clinical Data Warehouse (CDW) were used for reference.

3. Methods

3.1. The reference study

The reference study, Fibstroke [10], was a retrospective cardiovascular register study evaluating the associations of AF with stroke, TIA and intracranial hemorrhage. One of the aims of the Fibstroke study was to identify the incidence and timing of intracranial vascular events in relation to diagnosis and anticoagulation treatment of AF. Strokes, TIA events and intracranial bleeding events are collectively referred to as 'Index events' in this text.

The IT department of Turku University Hospital provided a data listing of the patients with AF before 2013 and at least one incidence of an Index event during 2003–2012. A comprehensive list of the ICD-10 codes used for the pre-screening is provided in Table 1. The list of patients was examined by a group of researchers, who manually reviewed the EHRs of potential study participants, one at a time, and verified their eligibility for the study. After exclusion of uncertain and non-eligible cases, they identified 2166 eligible patients matching the selection criteria.

Table 1
The eligibility criteria for the patients in the reference study (Fibstroke). ICD10 codes were used in all searches.

Criteria for initial screening	
1.	Disturbances in cerebral blood flow at any point during 10-year period, 2003–2012: I60.0-I60.9, I61.0 - I61.9, I62.0 - I62.9, I63.0 - I63.9, I64.0-I64.9, I65.0-I65.9, I66.0-I66.9, I69.0 - I69.9 or G45.0 - G45.9, G46.0 - G46.9 or S06.0- S06.9
2.	Had ever been diagnosed with atrial fibrillation or atrial flutter (AF), I48
Inclusion criteria	
1.	Stroke, transient ischemic attack (TIA), intracranial bleeding during 2003–2012
2.	AF
Exclusion criteria	
1.	Intracranial bleeding (S06) diagnosed before AF
2.	Post-operative AF only related to cardiac surgery procedure [a]
3.	Suspected TIA (G45) but not confirmed by neurologist [a]
4.	Diagnosis of transient global amnesia (G45.4) without evidence of cerebrovascular event
5.	Patients with data not available electronically
6.	Patients living in the catchment area for less than a year after the Index event[a]

[a] The criterion was not applicable when queried with the InSite EHR Research Platform and with Turku University Hospital's Clinical Data Warehouse query tool.

3.2. EHR research platform, InSite

InSite is a platform integrated with Turku University Hospital's CDW. It allows researchers to interact with an anonymized copy of the hospital's EHR system for validating and optimizing clinical trial protocols and for accelerating the recruitment of trial participants [11]. Data processing is performed under the hospital's control and the users of the EHR Research Platform can only see aggregated results, i.e. patient counts.

The InSite platform was used in the spring of 2018 according to the participant eligibility criteria of the reference study after formalizing the criteria into structured items. An example of the transcription of the selection criteria to searchable items is presented in Appendix 1. Three exclusion criteria, 'post-operative AF only related to cardiac surgery', 'suspected TIA but not confirmed by a neurologist' and 'patients living within the South-West Finland Hospital District for less than a year after the Index event' were not applicable. These criteria were not translatable into a structured format, or were not recorded in the EHR system.

3.3. Clinical Data Warehouse of Turku University Hospital

A hospital's EHR contains various kinds of patient data both in structured and unstructured format located in different data repositories of the hospital. Data lakes are designed to collect and combine large amounts of data located in different data sources and in several different native formats. The Clinical Data Warehouse, CDW, of Turku University Hospital is a structured repository on top of the data lake, and is already processed for data extraction and therefore optimal for research purposes. Data lakes with their data warehouses provide favorable infrastructure for integration of the EHR Research Platform to the use of a hospital's EHR data for research purposes. The CDW processes have been described previously and are available at GitHub [12].

Turku University Hospital's CDW was used in this study to verify whether the EHR Research Platform could identify, on the whole, the patients in the hospital's CDW, and, in addition, to compare the query results of the manual and the EHR Research Platform methods on a patient identifier level.

The CDW queries were completed during the summer and autumn of 2019 with the eligibility criteria of the reference study (Table 1). An example of the formalization of the criteria for the CDW query is presented in Appendix 2.

3.4. Comparison of patient counts

Three mutually exclusive categories were formed to explore the differences between the results provided by the search methods: 'Patients identified with CDW, but not with the manual search', 'patients not identified with CDW, but identified with the manual search' and 'patients identified with both methods'. Up to 50 examples of individual EHRs per each category were examined to define the basis on which they were or were not collected into the categories.

3.5. Time constraint analysis

The temporal relation between the two diagnoses was assessed because of its essential information value for searching of study subjects and also because of the complexity of formalizing an adequate time constraint query in a structured, digital format [13].

The diagnosis codes have been applicable in structured format at Turku University Hospital since 2004. Therefore, the time constraint testing was only performed within the patient population with AF diagnosed after that, leading to an analysis of 1002 patients's records.

Three groups were formed: 'patients with their first incidence of AF before the Index event', 'patients with their first incidence of AF after the Index event' and 'patients with their first incidence of AF and Index event at the same time'. As in the reference study, the time window was one day, i.e. the first AF occurred one day before or one day after the first diagnosis of the Index event. The records of 50 patients per category were reviewed.

3.6. Statistical analyses

Categorical data were summarized as patient counts and percentages. This was a descriptive study. No further statistical analyses were performed due to the descriptive nature of the study.

4. Results

The EHR Research Platform identified 5859 eligible patients, whereas the patient count with the manual search was 2166 (Table 2). The CDW search identified 5840 patients.

The EHR Research Platform and the CDW query apply the same search logic. This can also be seen in the proportions of different intracranial vascular incidents (stroke, TIA and intracranial bleeding) within the patient cohorts. For example, with both the EHR Research Platform and the CDW query, 82% of the patients had stroke as their Index event and 24% of them had TIAs (Table 2). Even if the patient count obtained with the manual search was much smaller than that obtained with the EHR Research platform, the corresponding proportions of intracranial vascular incidents in the manual search cohort were in adequate agreement from a clinical trial perspective.

4.1. Comparison of patient counts

4.1.1. Patients identified with CDWquery, but not with the manual search

Among the 5840 patients identified with the CDW query, there were 3674 patients not identified by the manual search (Table 3). The CDW query was not able to exclude a number of patients due to the unformalized exclusion criteria and due to its inability to distinguish between suspected and confirmed diagnoses in the structured data. For example, the diagnosis of AF or the Index event could be preliminary, unconfirmed or only suspected, as described in the free text field of a patient's EHR. All such patients were excluded in the manual search but were included in the CDW query.

4.1.2. Patients not identified with the CDW query, but found with the manual search

There were 133 patients not identified with the CDW query, but found with the manual search (Table 3). Based on an analysis of the records of 50 individual patients belonging to this category, this was due to two reasons: 41 of these 50 patients reviewed had the Index event during the years 2003–2004. As explained in section 3.5, the structured format of all diagnoses was implemented at Turku University Hospital in 2004, and consequently the CDW query could not find these patients. Secondly, the CDW query had correctly excluded another 9 of these 50 patients as they had an exclusion criterion ('Intracranial bleeding diagnosed before AF') recorded in the EHR, but for unknown reasons they had not been excluded from the manual search.

4.1.3. Patients found with both methods

There were 2033 patients identified with both the CDW query and the manual search. The identity of those patients was confirmed to be the same in both searches. Considering the reasons why the CDW query omitted 133 of the patients identified by the manual search (section 4.1.2.), it can be stated that the CDW query accurately identified all patients that had been correctly identified in the reference study.

4.2. Time constraint analysis

A proportion of patients was categorized differently into the three temporal categories (AF diagnosed before, after or at the same time with the Index event) by the CDW query and the manual search (Table 4). This was due to the different search logics and missing information in the structured data. The manual search collected the actual diagnosis dates for both AF and the Index event, using unstructured data. For the structured data in the EHR, as used by the CDW query, the diagnosis

Table 4
Patient population cohort with AF and index event divided into three categories based on the temporal relations of these two diagnoses. The search results were different due to a different search logic and an inappropriately narrow time window (one day) for these diagnoses.

	Patients with AF at least 1 day *before* Index event	Patients with AF *at the same time* with Index event	Patients with AF diagnosed at least 1 day *after* Index event
Manual search (n = 1002)	533 (53%)	270 (27%)	199 (20%)
Clinical Data Warehouse query (n = 1002)	412 (41%)	283 (28%)	307 (31%)

Table 2
Number of patients found in Turku University Hospital's EHR system as queried with a manual search, with the EHR Research Platform and with the hospital's Clinical Data Warehouse (CDW) query tool. *The sum of proportions of the index events within the cohort exceeds 100% as patients could have multiple events.*

	Manual search (Fibstroke)	EHR Research Platform (InSite)	CDW query
Patients with diagnosis of atrial fibrillation (AF) any time before 2013 and Index event in 2003–2012	2166	5859	5840
Composition of intracranial vascular incidents within the cohort:			
Patients with AF and stroke	1755 (81%)	4807 (82%)	4806 (82%)
Patients with AF and TIA (transient ischemic attack)	428 (20%)	1389 (24%)	1396 (24%)
Patients with AF and intracranial bleeding	313 (14%)	625 (11%)	596 (10%)

Table 3
Patient counts obtained with the CDW query compared to the manual search.

	Patients identified with CDW query (n = 5840)	
Patients identified with manual search (n = 2166)	Yes	No
Yes	2033 (35%)	133 (2%)
No	3674 (63%)	–

codes for each hospital stay were fixed to the date of the patient's discharge from the hospital, even if many of the diagnoses were established earlier during the hospital stay in question. This clearly impaired the precision of the diagnosis dates. Several cases were also identified where a diagnosis of AF was only added into the structured EHR data at the time of hospitalization because of the Index event, although a review of the unstructured data revealed that AF had already been diagnosed earlier.

5. Discussion

In this study, we examined the potential of an EHR Research Platform (InSite) for identifying the corresponding patients as found with a traditional manual search from the EHR system of Turku University Hospital. The manual search found 2166 eligible patients, whereas the EHR Research Platform identified 5859 patients. The CDW query of the hospital's EHR system, used as a reference for InSite because of its similar search logic and reliance on structured data, identified 5840 patients.

The patient counts were very similar in all tests performed with the InSite EHR Research Platform and with the hospital's own CDW query tool, indicating good technical performance of the InSite EHR Research Platform and an overall ability to identify the correct patients in the hospital's EHR system. Similar findings have also been reported in earlier studies [9].

Patient counts with the EHR Research Platform were 2.7 times higher than those obtained manually. We confirmed that the InSite EHR Research Platform can find the patients identified by conventional manual chart review. However, on top of that, it also included many additional patients in the patient counts. The additional patients identified with the EHR Research Platform mainly resulted from such criteria that could only be confirmed in the free text fields of the EHR system, i.e. in the unstructured data. Examples of such items in this study were diagnoses confirmed by certain specialists (exclusion criterion 3: Suspected TIA but not confirmed by neurologist), diagnoses related to a certain treatment (exclusion criterion 2: Post-operative AF only related to cardiac surgery procedure) and whether or not the diagnosis was placed as a tentative "working diagnosis".

In this study, five out of the eight eligibility criteria (63%) were possible to translate into structured, digital queries. According to Claerhout et al. [14], a median of 55% (38–89%) of the eligibility criteria in the trials covered by their analysis could be formalized into structured format for EHR research tools. The unformalized exclusion criteria led to excess numbers of patients in the patient cohort identified with the EHR query. It should also be noted that even if one would be able to formalize the selection criteria to digital queries, the EHR Research Platform did not discriminate between firmly established and tentative working diagnoses, which significantly increased the number of additional patients identified by the EHR query.

The EHR Research Platform seemed to accurately reflect the structured patient data available in the hospital's EHR system, but with more criteria left unstructured the more inaccurate it became, i.e. yielding too high estimates of the actual patient counts.

These limitations should be taken into account when using automated EHR Research tools for searching potential participants for a clinical trial. Still, already with the current potential of today's EHR Research Platforms, researchers can downsize the pool of potentially eligible patients. If the EHR query cannot be formulated to precisely reflect the original patient selection criteria, a manual review is needed after the EHR query. However, the workload of the manual selection process will be significantly decreased.

The variable quality and amount of data elements available in structured format at different hospitals [15] may hamper the comparability of patient counts between hospitals. Knowing this, hospitals have an incentive to further increase the quality of their EHR data in order to provide reliable estimates of their potential research participants. There are local and national regulations on the quality and characteristics of EHR systems in general, but there is a lack of unified descriptions for EHR data, and the functionality requirements of EHR systems are not enforced in a systematic manner [16].

5.1. Comparison of patient counts

The EHR Research Platform utilizes only structured data, whereas a manual query is capable of addressing all EHR data including the free text fields. In the manual search, only the diagnoses considered confirmed were included and patients with suspected events were discarded. With the EHR Research Platform, the diagnoses of AF or the Index events were collected from the structured data also if they were only suspected events; this difference could only be observed from the free text fields. This is a common challenge with EHR search tools utilizing only structured data; some crucial information may be missed if only structured data present in the EHR is used. Also, suspected diagnoses will remain in the EHR even if the final confirmed diagnosis is changed.

Another example of the restrictions faced by only using structured data are the AF events that occurred before the implementation of the structured format of the EHR system. When performing a search, the researcher should be aware of how old information can be reliably used. In addition, AF is commonly diagnosed in primary care, and is only recorded as a diagnosis in the hospital's EHR system when the patient is treated in the hospital. To overcome these limitations, data mining of the free-text fields is needed in order to increase the accuracy of patient identification.

5.2. Time constraint analysis

An eligibility criterion containing a time constraint requirement between two events, "X happened before Y", is rather common in clinical trials. As such a search criterion may be challenging to construct with an EHR Research Platform, it was an important item to investigate. When testing the temporal relations of diagnoses, an AF diagnosis could often be found in the unstructured data, which explains why the CDW query was not able find it and allocated fewer patients to the group "AF before Index event" than the manual search. Also, a different query logic was detected, which may explain why the CDW search allocated 13 more patients to the group "AF and Index event at the same time" than the manual search (Table 4): The CDW query collected the diagnosis date based on the patient's discharge date from the hospital, whereas the manual search collected the actual diagnosis dates of the events. Thus, shorter hospital stays are likely to provide more accurate results on the date of diagnosis than longer hospital stays. Due to the different query logic, it is not feasible to address the temporal relations between the two diagnoses with high temporal resolution, employing the current set-up of EHR data entry. Employing a time window of one week, instead of one day, would to some extent help to overcome this challenge, but would not eliminate it, as pre-existing AF may have been diagnosed in primary care, outside of the hospital, with no mention in the hospital's EHR system.

In this study, we presumed that the time constraint results obtained with the EHR Research Platform would be in line with those obtained with the CDW query, because the other test results were almost identical between these two methods. It also became clear that building a time constraint query required more expertise than the other queries, but after appropriate training the query building algorithms were assimilated [13].

5.3. Limitations

Currently, the EHR Research Platform only identifies structured data. Increasing amounts of data are being rendered to structured format, but currently, this is a limiting factor in the use of EHR data, as

some essential medical information remains to be only reported in the free text fields of EHR systems.

Despite some attempts to harmonize the creation of patient selection criteria in clinical trial protocols [14], there is no common standard available for formalizing the criteria presented in written clinical trial protocols into computable items. In this study, formalization of the criteria queried with the EHR Research Platform was undertaken by one person (NL) and formalization to CDW queries by another person (JMV). In spite of that similar patient counts were reached. By using this arrangement, we aimed at increasing the validity of the formalization process.

Due to additional procedural steps taken in the initial phase of the manual search process, the raw data listing used in the reference study was slightly different from the data source used in the CDW and EHR queries, which may have caused some variation in the results. As we were able to identify the patients in the samples reviewed and in the overlapping cohort and to confirm that both methods included the same patients, we find this limitation minor. Both the manual search and the EHR Research platform query (as well as the CDW query) used the same data lake of Turku University Hospital as the source data for the search.

We understand that the reference study was a retrospective register study, not aimed at recruiting clinical trial participants. However, it is a typical example of a procedure where patient records are manually searched for participant identification into a clinical trial.

The reference study included patients with suspected medical emergencies, such as a stroke, a TIA or an intracranial bleeding, hence requiring emergency care. Frequently, the suspected diagnoses recorded in emergency care became more exact or were changed during the hospital stay. As the EHR Research Platform was not capable of distinguishing between suspected or tentative working diagnoses and confirmed diagnoses, the result is differing patient counts. In a different reference study conducted with patients not in need of urgent medical treatment or without unconfirmed diagnoses, the number of suspected diagnoses would have had less influence on the patient counts. Therefore, considering the current capabilities of EHR Research tools, they perform their best in trials searching for patients with chronic diseases or other already confirmed conditions.

6. Conclusions

When querying for potential trial participants with a set of eligibility criteria, we conclude that an automated EHR research platform may be a useful tool for this purpose. Notably, though, when an important criterion, such as a diagnosis, a feature of a medication or a laboratory examination result is not available in structured format, different query search logics and criteria not translatable to structured digital queries will lead to discrepancies in patient counts compared to a conventional manual search. Researchers need to try to correctly formalize their selection criteria to digitally programmable items, understanding how the structured digital data represent and reflect those criteria. They also need to evaluate the quality of the structured digital data available in their hospital's EHR system. After taking this into account, the EHR Research Platform appears to be a feasible tool with potential for identifying eligible patients for clinical trials, or enabling to limit the population of possibly eligible patients and decreasing the need for manual review work. Major emphasis needs to be placed on describing the inclusion and exclusion criteria of the study protocol in a structured manner, on increasing the amount and quality of structured data in the EHR systems used by hospitals and on developing the use of unstructured EHR data. We believe that the use of EHR Research Platforms for this purpose will increase in the near future.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements

The authors would like to thank Eliisa Löyttyniemi (Department of Biostatistics, University of Turku, Finland) for expert advice in statistical matters.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.conctc.2020.100692.

References

[1] J.M. Foster, S.M. Sawyer, L. Smith, H.K. Reddel, T. Usherwood, Barriers and Facilitators to Patient Recruitment to a Cluster Randomized Controlled Trial in Primary Care: Lessons for Future Trials, 2011, https://doi.org/10.1186/s12874-015-0012-3.

[2] L.C. Hanson, J. Bull, K. Wessell, L. Massie, R.E. Bennett, J.S. Kutner, N.M. Aziz, A. Abernethy, Strategies to support recruitment of patients with life-limiting illness for research: the Palliative Care Research Cooperative Group, J. Pain Symptom Manag. 48 (2014) 1021–1030, https://doi.org/10.1016/j.jpainsymman.2014.04.008.

[3] A.M. McDonald, R.C. Knight, M.K. Campbell, V.A. Entwistle, A.M. Grant, J.A. Cook, D.R. Elbourne, D. Francis, J. Garcia, I. Roberts, C. Snowdon, What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies, Trials 7 (2006) 9, https://doi.org/10.1186/1745-6215-7-9.

[4] D.R. Kitterman, S.K. Cheng, D.M. Dilts, E.S. Orwoll, The prevalence and economic impact of low-enrolling clinical studies at an academic medical center, Acad. Med. 86 (2011) 1360–1366, https://doi.org/10.1097/ACM.0b013e3182306440.

[5] P. Bower, S. Wilson, N. Mathers, How Often Do UK Primary Care Trials Face Recruitment Delays?, 2007, https://doi.org/10.1093/fampra/cmm051.

[6] E. Kellar, S. Bornstein, A. Caban, M. Crouthamel, C. Celingant, P.A. McIntire, C. Johnson, P. Mehta, V. Sikirica, B. Wilson, Optimizing the use of electronic data sources in clinical trials, Ther. Innov. Regul. Sci. 51 (2017) 551–567, https://doi.org/10.1177/2168479017718875.

[7] Y.S. Lai, J.D. Afseth, A review of the impact of utilising electronic medical records for clinical research recruitment, Clin. Trials (2019), https://doi.org/10.1177/1740774519829709, 174077451982970.

[8] Us Food and Drug Administration, Use of Electronic Health Record Data in Clinical Investigations Guidance for Industry, 2018 accessed, https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htmand/orhttps://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/default.htmand/or. (Accessed 29 April 2019).

[9] J. Doods, R. Bache, M. McGilchrist, C. Daniel, M. Dugas, F. Fritz, Piloting the EHR4CR feasibility platform across Europe, Methods Inf. Med. 53 (2014) 264–268, https://doi.org/10.3414/ME13-01-0134.

[10] A. Palomäki, REAL-LIFE CHALLENGES of STROKE PREVENTION IN PATIENTS with ATRIAL FIBRILLATION – the FibStroke Study, 2017, ISBN 978-951-29-6775-9. http://urn.fi/URN.

[11] (n.d. InSite, accessed, https://www.insiteplatform.com/. (Accessed 6 March 2020).

[12] data-refinery-book/ACI_book-2018-09-24.pdf at master.Auriadih/data-Refinery-Book GitHub, (n.d.), accessed, https://github.com/auriadih/data-refinery-book/blob/master/ACI_book-2018-09-24.pdf. (Accessed 15 April 2020).

[13] I. Soto-Rey, A. N'Dja, J. Cunningham, A. Newe, B. Trinczek, C. Lafitte, B. Sedlmayr, F. Fritz, User satisfaction evaluation of the EHR4CR query builder: a multisite patient count cohort system, BioMed Res. Int. 2015 (2015) 801436, https://doi.org/10.1155/2015/801436.

[14] B. Claerhout, D. Kalra, C. Mueller, G. Singh, N. Ammour, L. Meloni, J. Blomster, M. Hopley, G. Kafatos, A. Garvey, P. Kuhn, M. Lewi, B. Vannieuwenhuyse, B. Marchal, K. Patel, C. Schindler, Federated Electronic Health Records Research Technology to Support Clinical Trial Protocol Optimization: Evidence from EHR4CR and the InSite Platform, 2019, https://doi.org/10.1016/j.jbi.2018.12.004.

[15] J. Doods, C. Lafitte, N. Ulliac-Sagnes, J. Proeve, F. Botteri, R. Walls, A. Sykes, M. Dugas, F. Fritz, A European inventory of data elements for patient recruitment, Stud. Health Technol. Inf. 210 (2015) 506–510. https://www-ncbi-nlm-nih-gov.ezproxy.utu.fi/.

[16] N.G. Weiskopf, C. Weng, Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research, J. Am. Med. Inf. Assoc. 20 (2013) 144–151, https://doi.org/10.1136/amiajnl-2011-000681.