Journal of Biomedical Informatics 148 (2023) 104553

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Special Communication

Electronic health records (EHRs) in clinical research and platform trials: Application of the innovative EHR-based methods developed by EU-PEARL

Giulia Lombardo [a], Camille Couvert [b], Melisa Kose [a], Amina Begum [a], Cecile Spiertz [c], Courtney Worrell [a], Danny Hasselbaink [d], Eva-Maria Didden [e], Luca Sforzini [a], Marija Todorovic [f], Martine Lewi [g], Mollie Brown [a], Morgan Vaterkowski [h], Nancy Gullet [a], Nare Amasi-Hartoonian [a], Nicolas Griffon [i,j], Raluca Pais [k], Sarai Rodriguez Navarro [l], Andreas Kremer [m], Christophe Maes [n], Eng Hooi Tan [o], Maxim Moinat [p], Joan Genescà Ferrer [l], Carmine M. Pariante [a], Dipak Kalra [q], Nadir Ammour [b], Susana Kalko [l,*]

[a] King's College London, Institute of Psychiatry, Psychology and Neuroscience, Department of Psychological Medicine, London, UK
[b] Sanofi R&D, Global Development, Clinical Science & Operations, Chilly-Mazarin, France
[c] The Janssen Pharmaceutical Companies of Johnson & Johnson, Leiden, The Netherlands
[d] Janssen Biologics B.V., Leiden, the Netherlands
[e] Actelion, a Janssen company of Johnson & Johnson, Allschwil, Basel-Country, Switzerland
[f] Johnson & Johnson Clinical Operations (JJCO), Johnson & Johnson company, Belgrade, Serbia
[g] Global Commercial Strategy Organization, the Janssen Pharmaceutical Companies of Johnson & Johnson, Raritan, New Jersey, USA
[h] Assistance Publique Hôpitaux de Paris, IT Department, Innovation and Data, Paris, France, and EPITA EPITA School of Engineering and Computer Science, Paris, France
[i] Information Technology Department, AP-HP, Paris, France
[j] LIMICS, Inserm U1142, Sorbonne Université, Paris, France
[k] Sorbonne Université, Assistance Publique-Hôpitaux de Paris, Hôpital Pitié-Salpètrière, Institute of Cardiometabolism and Nutrition, INSERM UMRS_938, Paris, France
[l] Vall d'Hebron Research Institute (VHIR), Barcelona, Spain
[m] Information Technology for Translational Medicine, ITTM S.A, House of BioHealth, Esch-sur-Alzette, Luxembourg
[n] The European Institute for Innovation through health data, and Department Public Health and Primary Care, Unit of Medical Informatics and Statistics, Faculty of Medicine and Health Sciences, Ghent University, Gent, Belgium
[o] Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK
[p] Erasmus University Medical Center, Rotterdam, the Netherlands
[q] The European Institute for Innovation through Health Data and Visiting Professor, University of Ghent, Gent, Belgium

ARTICLE INFO

Keywords:
Electronic Health Records
Clinical research
Platform Trials
EU-PEARL

ABSTRACT

Objective: Electronic Health Record (EHR) systems are digital platforms in clinical practice used to collect patients' clinical information related to their health status and represents a useful storage of real-world data. EHRs have a potential role in research studies, in particular, in platform trials. Platform trials are innovative trial designs including multiple trial arms (conducted simultaneously and/or sequentially) on different treatments under a single master protocol. However, the use of EHRs in research comes with important challenges such as incompleteness of records and the need to translate trial eligibility criteria into interoperable queries. In this paper, we aim to review and to describe our proposed innovative methods to tackle some of the most important challenges identified. This work is part of the Innovative Medicines Initiative (IMI) EU Patient-cEntric clinicAl tRial pLatforms (EU-PEARL) project's work package 3 (WP3), whose objective is to deliver tools and guidance for EHR-based protocol feasibility assessment, clinical site selection, and patient pre-screening in platform trials, investing in the building of a data-driven clinical network framework that can execute these complex innovative designs for which feasibility assessments are critically important.

Methods: ISO standards and relevant references informed a readiness survey, producing 354 criteria with corresponding questions selected and harmonised through a 7-round scoring process (0–1) in stakeholder meetings, with 85% of consensus being the threshold of acceptance for a criterium/question. ATLAS cohort definition and

* Corresponding author.
E-mail addresses: giulia.lombardo@kcl.ac.uk (G. Lombardo), kalko.susana@gmail.com (S. Kalko).

https://doi.org/10.1016/j.jbi.2023.104553
Received 28 March 2023; Received in revised form 13 November 2023; Accepted 20 November 2023
Available online 22 November 2023

Cohort Diagnostics were mainly used to create the trial feasibility eligibility (I/E) criteria as executable interoperable queries.
Results: The WP3/EU-PEARL group developed a readiness survey (eSurvey) for an efficient selection of clinical sites with suitable EHRs, consisting of yes-or-no questions, and a set-up of interoperable proxy queries using physicians' defined trial criteria. Both actions facilitate recruiting trial participants and alignment between study costs/timelines and data-driven recruitment potential.
Conclusion: The eSurvey will help create an archive of clinical sites with mature EHR systems suitable to participate in clinical trials/platform trials, and the interoperable proxy queries of trial eligibility criteria will help identify the number of potential participants. Ultimately, these tools will contribute to the production of EHR-based protocol design.

1. Electronic Health records and platform Trials: An overview

Electronic health record (EHR) systems allow the collection of patient-level real-world data (RWD) in the context of decision-making for patient clinical care (*US Department of Health and Human Services* [1]). However, their main domain has recently expanded and EHRs are now seen as an attractive tool in research.

EHRs store patient care data that could be leveraged in generating real-world evidence (RWE) and ultimately, help facilitate the design, planning, and conduct of large-scale and complex approaches, as in the case of platform trials, where a continuous influx of trial participants needs to be ensured. Platform trials have multiple trial arms with a single common master protocol, allowing the investigation of multiple targeted therapies, simultaneously or sequentially [2,3], in a continuous manner with trial treatments entering or leaving the study based on a decision algorithm [4,5].

One of the benefits of platform trials is that a lower proportion of all recruited trial participants are in a (shared) control arm or standard of care arm, and a greater number of trial participants have the potential to benefit from an innovative treatment. It is also possible to discontinue a study arm (e.g., due to safety issues or poor efficacy) and to transfer trial participants after a suitable washout period to a different intervention arm. Platform trial designs are therefore ideally suited to rare diseases, where patient numbers are small to start with. So far, successful platform trials have been reported in oncological research or, more recently, in coronavirus disease (COVID-19)-related research. The advantages of a well-established design in these fields can be leveraged in other research areas. The Investigation of Serial Studies to Predict Your Therapeutic Response through Imaging and Molecular Analysis 2 (I-SPY2-NCT01042379) is a platform trial investigating high-risk clinical stage II or III breast cancer, and it is a well-known representative of the potential of platform design in research. A mainstay of I-SPY2 is offering new tailored and targeted rapid treatments (or combination of treatments), thus enabling a prompt assessment of the efficacy for specific subtypes of trial participants based on predictors established in I-SPY1 [6,7,8]. Furthermore, the platform trial design has been recently adopted to evaluate potential COVID-19 therapies. A valuable review of the advantages of platform trials in COVID-19 research has been delivered by Stallard et al. [9]. The authors described the advantages of having an adaptive design available when information such as the progress of a disease and the features of patients is uncertain. Valuable examples are the Randomised Evaluation of COVID-19 Therapy (RECOVERY) study and the Platform Randomised Trial of Treatments in the Community for Pandemic and Epidemic Illnesses (PRINCIPLE) study [10,11].

Platform trials may contain a large observational cohort fulfilling a set of eligibility criteria documented within a master protocol for that disease area. Each trial compound may introduce a refinement to the generic eligibility criteria, documented within supplementary protocols annexed to the master protocol. The concept of platform trials is extremely relevant nowadays, as precision medicine is becoming the preferred approach, in both research and clinical practice. This study design provides the opportunity to be adaptable in terms of replacing interventions based on the effectiveness of the study arms, with an enhanced prospect of success. The intervention arms in a platform trial do not necessarily have to run synchronously: new products, and therefore new intervention arms, can be added at any point and populated with trial participants who are still in the longitudinal observation cohort who are not already enrolled in a different intervention arm.

1.1. Identifying challenges and current objectives

Even though EHRs might be significant assets in large-scale approaches, their use is not straightforward, and some challenges need to be considered. Overall, as highlighted by Corrigan-Curay et al. [12], EHRs have not been primarily designed to be used in clinical research and thus, they often present data organised in a non-research-friendly way, with important inconsistencies throughout the systems [12]. Therefore, EHRs need to be appropriately and correctly implemented to be able to be used in support of potential participants pre-screening in clinical trials and platform trials. Hence, the main aim of this paper is to discuss the use of EHR data in research. Specifically, we aim to discuss their use in clinical platform trials, by describing the challenges and the application of innovative EHR-based methods to improve the planning and execution of platform trials.

2. EU-PEARL project and work package 3

2.1. EU-PEARL Project: An overview

The innovative EHR-based methods have been developed as part of the Innovative Medicines Initiative (IMI) EU Patient-cEntric clinicAl tRial pLatforms (EU-PEARL, http://eu-pearl.eu). As described by Sforzini et al. [2], EU-PEARL is a European project funded by the IMI aiming to create a design of an integrated research platform (IRP). The project aims to develop methodologies and tools to support the large-scale uptake of platform trial designs across Europe, and potentially globally. It aims to develop generic approaches but focuses on four major clinical areas of high public health relevance: major depressive disorder (MDD), tuberculosis (TB), non-alcoholic steatohepatitis (NASH), and neurofibromatosis (NF) [13]. During the EU-PEARL General Assembly Meeting, representatives from other well-known running platform trials have been invited to share their experience in setting up their own research platforms, by sharing lessons they learned and the challenges they faced (A Randomised, Embedded, Multi-factorial, Adaptive Platform Trial for Community-Acquired Pneumonia (REMAP-CAP) [14], the HEALY Amyotrophic Lateral Sclerosis (ALS) platform trial [15], and Beat Acute Myeloid Leukemia (AML) platform trial [16]. The Leukaemia and Lymphoma Society leading the Beat AML platform has shared how they are successfully leveraging the data from the EHR to screen trial participants and direct them to the best available sub-study based on their specific information. One of the major advantages of this European project is the presence of various stakeholders within this reality: clinicians, patient representatives, clinical sites, regulators, solution providers, and pharmaceutical companies, with Janssen and Sanofi co-authors representing their perspectives. There is an important cooperation between all work packages (WPs) and, as highlighted by Cowie et al. [17], the collaboration between different perspectives is an essential step towards the effective use of EHRs in clinical research.

Furthermore, a master protocol in each disease-specific area is being developed and may be reported in future publications.

2.2. Work package 3: Specific aims

Within the EU-PEARL project, WP3 is designing an innovative method to inform the protocol designers for the selection of clinical sites through evidence of existing patients matching protocol eligibility criteria in high-quality EHR systems. WP3 is developing a generic process for interrogating EHRs to accelerate the processes of site selection and potential participant recruitment, in compliance with the General Data Protection Regulation (GDPR) requirements for the protection of personal data (679/2016) [18]. The outcomes would allow querying of EHR data for healthcare organisations not connected to a central federated data platform, mainly using the Observational Medical Outcomes Partnership Common Data model [19] (OMOP CDM) to translate the master protocols eligibility criteria into computable elements. OMOP CDM is a central component of The Observational Health Data Sciences and Informatics (OHDSI) program, a multi-stakeholder, interdisciplinary, and collaborative ecosystem aimed to bring out the value of health data through large-scale analytics [20]. In OHDSI, all solutions are open-source. This would make it easy to be run against hospital EHRs already mapped to the OMOP data model, as well as against hospital EHRs not yet mapped to the OMOP data model, facilitating the conversion of the queries to the hospital-specific EHR format. To be able to do so, as the hospitals are the main providers of data, it is essential to identify sites capable of using their EHRs for research, referred to as EHRs-enabled sites, from those who cannot.

WP3 has developed an EHR system and data readiness requirements survey (eSurvey) as a continuation of prior work in the Electronic Health Records to Electronic Data Capture systems (EHR2EDC) project [21]. This readiness assessment survey is for EHR-enabled site network membership, confirming through a formalised self-assessment process that their EHR systems can inform protocol design. See Table 1 for the Statement of Significance.

3. The value of hospital EHRs for the design of platform trial protocols

Clinical trials are the bedrock of evidence generation to demonstrate the efficacy and safety of a new medicinal product, in order to gain regulatory licensing approval, and EHRs can be utilised within research, as an extremely useful resource to improve potential participant selection and recruitment. One of the purposes of the EU-PEARL project is to address the limitations of the current state of EHRs in assisting recruitment, in particular in platform trials. Participant recruitment often faces difficulties in meeting the initial target sample size of participants with 50 % of trials failing to meet their expected progress [22] and a third of trials being terminated due to insufficient recruitment [23], possibly producing underpowered trials. Even though EHRs are not systematically recommended to perform the eligibility screening for candidate participants in research, when the EHR infrastructure is entrenched, the cost per participant might be lower than the cost per participant in the case of manual search [24]. This brings to light the potential cost-wise advantage of the use of EHRs in recruitment upon the presence of a well-established infrastructure [24].

Traditional recruitment estimations often rely on informal, simplistic assessments of participants with the condition, rather than precise eligibility criteria matching. Manual chart reviews, while accurate, are time-consuming and resource-intensive, posing challenges for most trial sites. Leveraging the computable data within the hospital EHRs has long been recognised as a desirable alternative. Implementing the search on EHRs with automated queries (i.e., using of informatics tools in the screening process of potential candidates) can drastically lower the amount of work to identify potential participants, as highlighted by Ni et al. [25]. The capability to embed feasibility queries into in-house EHR systems was demonstrated in the late 1990 s, for example, at the Royal Marsden Hospital in London (unpublished, on-site demonstration seen by author Kalra in 2000). Evidence of improved recruitment through EHR querying has also been well documented [26,27,28].

The first major generic methodology to implement this feasibility querying independently of the EHR system in use, including appropriate data protection, governance, interoperability specification, and Application Programming Interfaces (APIs), was published and piloted by the Electronic Health Records for Clinical Research project (EHR4CR) 2012–2015 [29]. The EHR4CR specifications have stimulated a growing market of health information and communication technology (ICT) products that can connect through a federated network to multiple hospital sites and cascade candidate protocol feasibility queries to obtain participant counts (e.g., The TriNetX Platform [30], The Clinerion query builder [31], ACT Network [32]). These counts, that provide an aggregated number of participants, can often be generated rapidly in near to real time and allow a protocol designer (e.g., within a pharmaceutical company) to iteratively adapt the eligibility criteria until there is a sufficient predicted number of potential participants to make the trial viable. Although every clinical trial has some unique criteria, there is a common core of eligibility criteria that are frequently recurring across protocols, across therapeutic areas, which makes the implementation of these federated query tools and the mapping of data from hospital EHR systems relatively tractable [33]. There has been more recent work to explore the possibility of incorporating protocol feasibility queries into the OHDSI [20] tools and network, which could greatly increase the scale of the available hospital sites whose data could be examined [34]. However, it should be noted that not all eligibility criteria are capable of being represented as computable query expressions as some require subjective interpretation by a domain expert [35]. Investment is needed to better educate clinical trial protocol designers about the kinds of data that are realistic to find in hospital EHRs, and how that information is structured and coded, in order to encourage a higher proportion of eligibility criteria to be capable of computable representation.

The case for using hospital EHRs to assess the feasibility of a platform trial and of each newly introduced intervention arm is even stronger for platform trials, where the query may need to match individual participants to more than one arm, to allow participants to be suitably distributed across study arms so that they are all viable.

One aspect of the EU-PEARL work plan is to examine the eligibility criteria being proposed for the four disease area master protocols in order to undertake some feasibility queries across the eight university hospitals that are part of the project consortium, and which are members of the European University Hospital Alliance (EUHA) (https://www.euhalliance.eu/). These pilot eligibility criteria will be modelled using the OHDSI tools (with support from the IMI EHDEN project [36]) and executed on exported data from each of the hospitals represented according to the OMOP CDM [19,37,38,39]. The OMOP CDM has been developed to do observational healthcare research at a large scale, from drug utilisation to comparative effectiveness studies. In Europe, the

Table 1
Statement of Significance.

Problem	Leveraging electronic health records (EHRs) in platform clinical trials could improve the pre-screening process for participant recruitment, but several issues undermine their use in research.
What is already known	The process to select clinical sites which have harmonised EHRs has not been created yet, and the protocol eligibility criteria are not always fit to allow an efficient pre-screening process.
What this paper adds	The EU Patient-cEntric clinicAl tRial pLatforms (EU-PEARL) project has developed a readiness survey for the selection of clinical sites and trial feasibility criteria as interoperable queries to obtain the number of potential participants at those sites.

adoption of the OMOP CDM has been driven by the IMI EHDEN project consortium with 166 data partners from 27 countries using the OMOP CDM. To support research using the OMOP CDM, there is a collection of open-source software maintained by the OHDSI community [20].

The use of hospital EHRs described here focuses on the design of a trial, in particular a platform trial. However, there has been additional work to demonstrate that the data within the EHR of a patient who has given consent to participate in a trial and given consent for the reuse of their data can be soundly transferred from EHR to electronic data capture (EDC) systems used in clinical trials. The EHR2EDC [22] approach, including its data mappings and governance model, is also progressively being picked up by ICT companies and will become an increasingly valuable solution to diminish the current practice of manual data transcription which is expensive and error-prone. EU-PEARL has benefited from the learning of EHR2EDC.

4. WP3/EU-PEARL innovative methods to implement the application of EHR in RCTs/Platform trials

4.1. Challenge #1: Readiness of EHRs

The use of EHR data in clinical research and platform trials is associated with various stakeholder benefits (*see* Table 2). However, some challenges need to be considered, such as the presence of non-research-friendly data [28], missing data, and discrepancies between study staff evaluation during computerised queries using EHR data. A large proportion of data might be recorded in unstructured form (e.g., medical notes, free-text summaries) [39] or recorded from historic data using a coding system that has not been mapped to standard concepts. These features can limit the EHR search and cause a discrepancy in the numbers estimated from a computerised search of structured data versus a manual review of all available information. A valid method to overcome the difficulties derived from the presence of unstructured data in EHRs is the use of Natural Language Processing (NLP) to detect relevant information from free-text sections. Meystre et al. [40] detected a key role of NLP in breast cancer clinical trials when considering the use of EHRs in extracting eligibility criteria and trial eligibility from clinical notes, making this method appealing to create an efficient screening process for trial participants [40]. NLP is indeed important to produce accurate queries crucial to understanding the process and its challenges. This service might be offered by commercial companies; however, we have not evaluated commercial offerings of NLP software in the context of this project. EHRs vary in complexity and readiness between organisations. To address this, there is a need for a tool to identify and select mature clinical sites with suitable EHR systems for platform trials, a goal that aligns with the aims of WP3/EU-PEARL. Once the EHR quality is confirmed, a second evaluation involves querying the system for potentially eligible patients for the IRP protocol using interoperable queries. In this way, researchers and clinicians can construct a network of clinical sites suitable for the development of complex clinical trials.

Table 2
Stakeholder benefits using EHR data in clinical research and platform trials.

Stakeholder Benefits
Patients & Ethics
• Secure an adequate recruitment within allocated timelines and budgets, given that exposing patients to non-conclusive research is unethical as it would not be a reliable representation of the population of interest.
Greater ease of determining patient eligibility.
Increased opportunities to address longstanding underrepresentation of certain groups in clinical trials and thus, to produce evidence that is more informative for a broader patient population.Offering a treatment option, when diagnosed with a medical condition with no satisfactory treatment, in the case of paediatric population and/or rare diseases (which are linked to slow recruitment rates and difficulties in meeting target size)
.
Clinicians
• Support in managing the increased complexity of patient recruitment e.g., complex inclusion and exclusion criteria.
Select and accept protocols based on available and sufficient number of eligible participants matching the study protocols eligibility criteria.
Make recruitment predictable with realistic recruitment targets based on RWD analysis.
Align study budgets and timelines with realistic data-driven recruitment potential.
Support the need for targeting undiagnosed patients that could be eligible for trials.
Support the use of clinical trials as a treatment option for diseases with high unmet needs.
Reduce workload and efforts to assess recruitment potential based on study eligibility criteria.
Speed up recruitment and access to patients beyond the individual practice level, instead enabling institution-wide recruitment.
Regulatory
• Obtain data from under-represented populations of patients.
Access evidence and data, in relation to the target patient population.
Industry
• Design protocols and eligibility criteria are real-world data-driven.Improve reliability of study placement and enrolment plans: (i) more accurate recruitment planning and prediction, (ii) reduction of non-enrolling and low recruiting clinical sites, (iii)
avoid zero enrolling study.
Accelerate the clinical trial recruitment process to provide faster access for patients to innovative treatments.

4.2. EU-PEARL/WP3: EHR readiness survey

As a first step, WP3 has developed an online tool to be able to assess the maturity, quality, and suitability of hospital EHR systems to be used for research purposes. This tool named eSurvey was built from several already published references, the Investigator Site eSource Readiness Assessment 2020 (eSRA) from eClinicalForum [41], the EHR2EDC readiness survey v1 [29], the TransCelerate eSource Informatics Information Reference Guide for Clinical Research sites 2019 [42], the European Medicines Agency Reflection paper on Good Clinical Practice (GCP) compliance in relation to trial master files for management, audit and inspection of clinical trials 2015 [43], and the Medicines & Healthcare products Regulatory Agency (MHRA) Position Statement and Guidance on Electronic Health Records 2015 [44]. Moreover, ISO standards have together informed the readiness survey (ISO 13606:2019 Electronic Health Record communication; ISO 18308: 2017 Requirements for an Electronic Health Record Architecture; ISO 13940: 2015 System of concepts to support continuity of care; ISO 21090: 2011 Harmonised data types for information interchange; ISO 27789: 2021 Audit trials for electronic health records). Two independent researchers screened all the criteria for duplicity from the referenced articles. This resulted in a set of 354 unique criteria. The criteria, the corresponding questions, and the defining of the subdomains were achieved in co-creation with all relevant stakeholders (public and private). In total 27 individuals (EHR, Clinical Trial, ICT, technical experts, data experts from hospitals, and sponsors) selected the criteria and the corresponding questions. The criteria and the corresponding questions were selected and harmonized during interactive workshops, online meetings, and 1-to-1 site meetings. The following methodology was used: each stakeholder had to score each question: score 1 (indicating a question to be kept) and score 0 (indicating a question that should be taken out of the questionnaire). The consensus was defined as follows: a question was accepted if at least 85 % of the stakeholders gave a score of 1. If less than 85 % was achieved for a particular question, a consensus had to be reached during a teleconference. After seven rounds, a consensus was reached regarding whether to accept a criterium/question. Once the criteria and corresponding questions were defined, the “must-have” and “desirable” criteria and questions were determined following the same methodology. After two rounds, all criteria and questions were categorised into either “must have”, the “desirable”, or the “optional” category. Items under the “desirable” category are those which are not mandatory, but it is preferable to retrieve a response; whereas items under the “optional” category imply that there is no consequence

incurred in the absence of a response.

This online tool consists of yes-or-no questions (N = 115, counting 25 "must-have" and 90 "desirable" items) divided into specific domains (N = 6), each one important in establishing the digital possibility of finding eligible participants and the data needed to comply with the trial protocol design. The six domains are the following: (i) general information related to the hospital site (e.g., the main point of contact and authorisation to share information); (ii) patient pre-identification to detect potential trial participants (i.e., unique patient identifier and detecting individuals participating in other studies); (iii) electronic data capture for studies (i.e., the digital connection between EHR and other internal and external patient data systems); (iv) EHR quality management, good clinical practice (GCP), and general data protection regulation (GDPR) (e.g., back-up, retention periods, revision cycles); (v) EHR2EDC component (e.g., coding and mapping systems, data format types); (vi) data quality management (e.g., on internal and external generated data, data quality requirements in service level agreements, data quality domains). The scoring system is as follows. Score 3 was given for questions that are mandatory to be answered with a yes for a site to qualify as a mature EHR site to participate in platform trials. Score 2 questions are desirable and score 1 questions are optional for a site. Given that each of the 115 questions has a score, it is possible to visualise per hospital site the overall score and the scores per domain. This allows for compiling a catalogue of EHR mature sites ready to participate in platform trials. By developing this tool, we would like to allow each hospital and site to be able to take this survey to self-assess their maturity in using their EHR for research, therefore enabling them to join the EHR-enabled sites network. This online questionnaire can also be used by the sites as a roadmap for future improvement of their EHR systems.

Validation of the survey was the main objective of piloting the eSurvey Readiness Survey (eSurvey) in nine partnering hospitals from Spain, United Kingdom, France, Belgium, Germany, Netherlands, and Austria. As the eSurvey Readiness survey has different categories of questions to be answered by different roles within the hospital sites, the main responder of each specific domain was identified. The hospitals had five months to complete the survey. Six hospitals completed the survey, one hospital only partially completed the survey, and two hospitals did not start the survey. The objectives of piloting the survey were: i) the validation of the criteria and corresponding questions, ii) checking the phrasing of the questions, iii) checking whether there were missing questions, iv) checking whether there were missing domains, and v) checking whether specific domains were addressed in dedicated functions by the correct persons.

It was not the intention to benchmark the hospital sites; however, as we could analyse the results of the completed surveys, it was possible to generate some general but very interesting conclusions in an anonymous way. The overall maturity score (% yes answers, indicating they are compliant with the criteria) of the six hospitals that completed surveys is 49 %, 18 %, 28 %, 31 %, 6 % and 55 % respectively, indicating a very low eSurvey maturity score. The scores for the mandatory criteria (% yes compliant answers to all mandatory criteria) are 60 %, 40 %, 45 %, 50 %, 10 %, 80 %.

4.3. Challenge #2: Trial eligibility criteria

It is essential to compare, and adjust if necessary, information contained in these records and the eligibility criteria, a procedure that may have severe limitations as not all eligibility criteria could always be reliably captured in the EHRs causing inconsistencies between a given trial and potentially eligible trial participants [45].

The EHR-driven phenotyping is based on the selection of trial participants through specific criteria [46], and the presence of missing data might further increase the risk of bias. Even within one site, there is huge heterogeneity in completeness, depending on patient handling; inpatients will have a lot more information than outpatients will, and the selection of participants through querying healthcare datasets can be associated with an over-representation of "less-healthy individuals" [47]. There is a need for EHR-friendly selection criteria in a structured layout [27], and to adopt the use of major eligibility criteria for the prescreening process (i.e., a small set of machine-readable standardised significant eligibility criteria) [48].

4.4. EU-PEARL/WP3: Building queries for protocol feasibility and participant counts

The first step to interrogate EHR-enabled clinical sites is to adapt trial eligibility criteria to increase the re-use of EHR data for research. The use of eligibility criteria in clinical trials refines the cohort of interest to better assess the effect of an intervention. Leveraging EHR data for feasibility counts during the trial protocol design process can enable the successful achievement of the recruitment target and can help to avoid protocol amendments due to unrealistic inclusion/exclusion (I/E) criteria. In EU-PEARL, one of the most important missions is to guarantee interoperability between clinical sites [49] and to replace manual processes with automatic ones, where possible, using informatics tools. This replacement would help avoid possible human errors, save time, and significantly improve overall efficiency,

This is a complex challenge, primarily because hospitals have implemented their EHRs in customised ways and using different EHR solutions (i.e., developers and vendors). In this situation, the best solution is to have the warehouse data of the eligible sites (selected through the use of the EHR readiness survey) translated to some CDM. In EU-PEARL, the OMOP CDM has been selected to facilitate interoperability between sites. OMOP is created for observational studies; thus, an innovative aspect here is to leverage this data model in the case of interventional studies. This data standard can also be used in the EU-PEARL context of site selection and clinical trial feasibility.

As described in the introduction, we focus on the calculation of the number of potential patient (NPP) counts matching the lists of I/E criteria for the four diseases under study, defined by the clinicians and corresponding to their IRP needs. Even in the case of all sites using OMOP CDM, we have made several local adjustments due to the complexity and heterogeneity of the EHR structures on each clinical site. This challenging situation prompted us to adopt an incremental strategy to assess the feasibility of any of the disease-specific protocols to build, export, execute, and validate the queries across multiple hospitals (see Fig. 1).

We used the ATLAS cohort definition module from OHDSI [21] to create the trial feasibility eligibility I/E criteria as executable queries which can be shared between sites. ATLAS facilitated the creation of cohorts by integrating vocabulary search terms in structured data such as medical conditions or exposure to medicines to define cohorts. Sites using the OMOP CDM can execute these queries directly on their data to retrieve patient counts. In addition to ATLAS, we evaluated Cohort Diagnostics [50], a software package written in the R platform [51] for combining results for all cohort definitions from different sites. Cohort Diagnostics is fully interoperable with ATLAS cohort definitions and data in the OMOP CDM.

When lacking OMOP databases at any site, alternative queries may be constructed in Structured Query Language (SQL) or other computational languages compatible with local data warehouses, and aligned, to the extent possible, with those queries performed in the ATLAS tool. However, even OMOP-structured databases might require a modification of the query scripts. The current heterogeneity of the databases at the different sites usually forces the development of specific additional "ad-hoc" tools or scripts to align the querying strategies. Informaticians from different areas (data architects, software developers, bioinformaticians) must have a constant dialogue to supervise these tasks. To reach this objective we introduced a previous step for all queries, constructing "pre-code" texts in collaboration with medical experts in plain language and understandable to the informatics actors that have direct contact with their own EHR system. We followed a similar

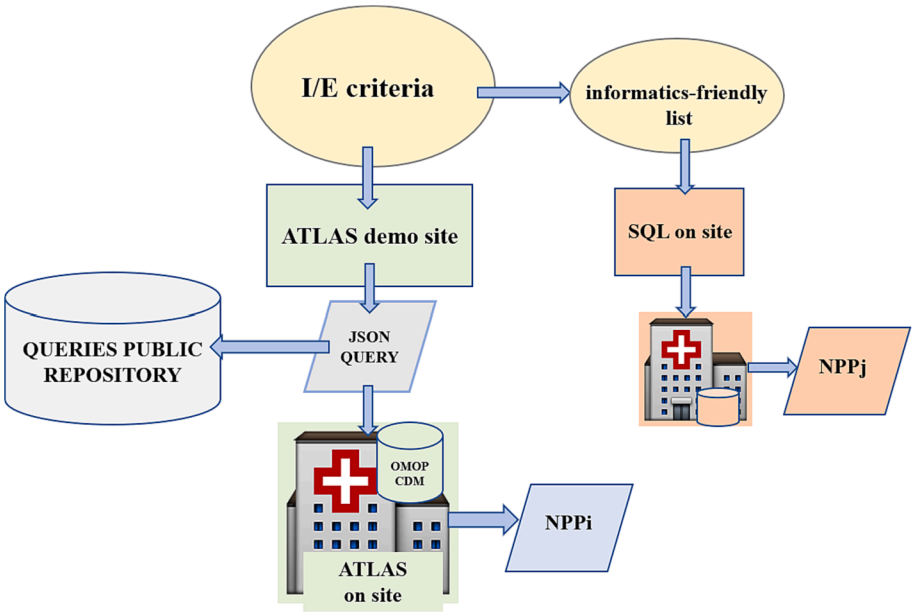

Fig. 1. Query procedures in the Main (OMOP) and Alternative (non-OMOP) roadmap Note: Input (I/E criteria defined by clinicians); outputs (Number of Potential Patients (NPP)). Close dialogue with informaticians or data scientist were necessary for Main or Alternative procedures. Main: in OHDSI-OMOP schema, we used the ATLAS Demo Site to derive JSON source code. Then, this JSON query was sent by email and was executed at VHIR and EMC respective OMOP databases, finally, each NPP was retrieved. An optimal reuse scheme is to have JSON queries at a repository open to the research community according to FAIR principles. Alternative: in a non-OMOP database, we had several rounds of discussions with local informaticians, and an "informatics-friendly" list of requirements was derived. Then, they executed their SQL code, highly dependent on the local EHR system. SLAM BRC was consulted in this way and its NPP was retrieved. The difference in set-up and turnaround time across centres varied from weeks to months between the main and alternative procedures.

sequence of statements and prioritisations of rules and lapse times, and we called them the "harmonised pre-code queries" (to be published), which may serve as guidance for future users.

Another important decision we made in the construction of the queries is that, in some cases, there is not sufficient or complete information in the warehouse databases to extract the NPP. Some diseases were described or codified recently, or there is an absence of defined "biomarkers" to guarantee the suspected diagnosis or that information is included in medical notes in free text (i.e., unstructured data) that cannot be captured in most of the cases. In these cases, the NPP may give zero counts, despite the clinicians knowing to have patients with those specific medical conditions. We should be prepared to further reformulate the query construction, changing or combining conditions to have results appropriate to known clinical prevalence. The results obtained using our approximate queries ("*proxy queries*") should be discussed with experienced clinicians and clinical teams to arrive at a consensus to select study participants, whether the calculated NPP matches the number of known patients in the specific disease areas, at a reasonable level of tolerance, and depending on the indication.

Our working group (WP3) has implemented the following process to translate trial eligibility criteria into computerised query criteria. For every step in the process, we involved clinicians from the Disease-Specific Work Packages (DSWP) to provide input on the definitions. We defined the anchor event, which is the first record of diagnosis for a disease-specific cohort over a specified observation period. Phenotypes were created for the disease conditions as these conditions may be recorded by more than one diagnosis code. Moreover, we further defined eligibility criteria like age and exclusion diagnoses with temporal reference to the date of the anchor event (e.g., no record of competing diagnosis in the two years prior to the inclusion diagnosis). These cohort definitions can be exported in different formats (e.g., SQL and JavaScript Object Notation (JSON)) which are interoperable with other sites using the OMOP CDM and ATLAS tools. Future publications with details on the creation and execution of queries, and results, are currently under production.

All the cohort definitions created in the EU-PEARL project will be shared in a publicly accessible repository, GitHub. GitHub (https://github.com) is a web service where version-controlled open-source software repositories can be hosted for collaboration. It can be freely downloaded for future users, who may use and change them to particular needs on their own sites.

As discussed, eligibility criteria are often documented in unstructured text, making it challenging for an automated cohort definition process. Translation of trial eligibility criteria into EHR phenotypes is a complex and time-consuming process [45]. Guiding antecedents on the translation of medical to informatics language may be found in the work of Boxwala et al. [52]. However, this work is focused on computer-based decision support (CDS) which does not align with the objectives of our work. Recent advancements in NLP and the adoption of CDMs have enabled cohort definitions to be executable and sharable across institutions more efficiently. The unstructured eligibility criteria of clinical trials can be translated into computable queries with a tool like Criteria2Query [53]. First, the information extraction pipeline parses eligibility criteria in free-text format into standardised data elements such as medical condition measurement or drug temporal and numerical values. Second, the query formulation pipeline generates automated cohort definitions. Lastly, this query may be exported to ATLAS which has a user-friendly web-based interface where manual review and editing can be performed prior to executing the query to obtain patient feasibility counts. Although the Criteria2Query approach is interesting, there are some limitations, including information extraction errors which would need a larger training data set for further improvements.

The WP3/EU-PEARL work will guide future studies in leveraging EHRs in clinical research. However, one of the limitations of the eSurvey is the difficulty in answering all the questions, as highlighted by several hospitals during a closing teleconference. This issue should be further investigated. In addition, there was no possibility to assess the sensitivity of the eSurvey and thus determine whether it can yield the desired patient cohort. An additional value of this manuscript is the elucidation of the methods and description of the newly developed tools. As future

directions, the use of an NLP tool is necessary to implement the screening process for trial participants. A desirable step forward is also the incorporation of genetic, genomic data, assessments from digital phenotyping tools, and patient-reported outcomes in EHRs. This will improve both patient monitoring and care, and characterisation of trial cohorts, offering the possibility of precision medicine with a deepened knowledge of the genetic nature of diseases [54,55,56].

5. Conclusion

The leveraging of EHR in clinical research and platform trials faces a fundamental step: to assess the maturity, quality, and suitability of hospital EHR systems to be used for research purposes. EU-PEARL developed a self-completion online survey to help classify clinical sites and identify which ones may be part of the clinical networks participating in the complex trials.

Next, it is relevant to highlight the challenge of the use of protocol eligibility criteria as an input to convert the aforementioned criteria into a standardised (here: OMOP CDM based) query. The criteria are originally designed with the aim of randomisation and recruitment, and they are based on the ability to perform the required investigations during the screening visit, independently of what is already documented in the study participants' EHRs. On the other hand, some criteria need to be added to generate results that are reflective of the number of potential candidates in a given time frame (i.e., how many patients currently treated with this condition meet these additional criteria), and thus, to estimate the most accurate number of eligible trial participants.

Finally, some key criteria might not be available for querying the hospital EHRs. Then the protocol criteria should be used to define some surrogate "proxy" criteria based on data more likely to be available in the EHRs. Ideally, the query should include some specific conditions allowing the count of patients actively seen at the institution during a specific period of time. EU-PEARL developed queries using I/E (or approximated I/E) criteria to extract the count number of potential participants in the complex trials. These queries were exported and used locally in the appropriate sites. Some of the queries developed will be made publicly accessible after the final acceptance of that publication.

In conclusion, the tools developed by WP3/EU-PEARL have the potential to leverage EHRs in complex clinical trials/platform trials with a focus on EHR-based protocol design and feasibility assessment, clinical site selection, and patient pre-screening for platform trials.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

"EU-PEARL has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 853966-2. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA and CHILDREN'S TUMOR FOUNDATION, GLOBAL ALLIANCE FOR TB DRUG DEVELOPMENT NON PROFIT ORGANISATION, SPRINGWORKS THERAPEUTICS INC."

References

[1] US Department of Health and Human Services: https://www.hhs.gov/.
[2] L. Sforzini, C. Worrell, M. Kose, I.M. Anderson, B. Aouizerate, V. Arolt, M. Bauer, B. T. Baune, P. Blier, A.J. Cleare, P.J. Cowen, T.G. Dinan, A. Fagiolini, I.N. Ferrier, U. Hegerl, A.D. Krystal, M. Leboyer, R.H. McAllister-Williams, R.S. McIntyre, A. Meyer-Lindenberg, A.H. Miller, C.B. Nemeroff, C. Normann, D. Nutt, S. Pallanti, L. Pani, B.W.J.H. Penninx, A.F. Schatzberg, R.C. Shelton, L.N. Yatham, A.H. Young, R. Zahn, G. Aislaitner, F. Butlen-Ducuing, C. Fletcher, M. Haberkamp, T. Laughren, F.L. Mäntylä, K. Schruers, A. Thomson, G. Arteaga-Henríquez, F. Benedetti, L. Cash-Gibson, W.R. Chae, H. De Smedt, S.M. Gold, W.J.G. Hoogendijk, V. J. Mondragón, E. Maron, J. Martynowicz, E. Melloni, C. Otte, G. Perez-Fuentes, S. Poletti, M.E. Schmidt, E. van de Ketterij, K. Woo, Y. Flossbach, J.A. Ramos-Quiroga, A.J. Savitz, C.M. Pariante, A Delphi-method-based consensus guideline for definition of treatment-resistant depression for clinical trials, Mol. Psychiatry 27 (3) (2022) 1286–1299, https://doi.org/10.1038/s41380-021-01381-x.
[3] J.J.H. Park, M.A. Detry, S. Murthy, G. Guyatt, E.J. Mills, How to Use and Interpret the Results of a Platform Trial: Users' Guide to the Medical Literature, JAMA 327 (1) (2022) 67–74, https://doi.org/10.1001/jama.2021.22507.
[4] J. Woodcock, L.M. LaVange, Master Protocols to Study Multiple Therapies, Multiple Diseases, or Both, N. Engl. J. Med. 377 (1) (2017) 62–70, https://doi.org/10.1056/NEJMra1510062.
[5] B.R. Saville, S.M. Berry, Efficiencies of platform clinical trials: A vision of the future, Clin Trials. 13 (3) (2016) 358–366, https://doi.org/10.1177/1740774515626362. Epub 2016 Feb 17 PMID: 26908536.
[6] C. Printz, I-SPY 2 may change how clinical trials are conducted: researchers aim to accelerate approvals of cancer drugs, Cancer 119 (11) (2013) 1925–1927, https://doi.org/10.1002/cncr.28172. PMID: 23695808.
[7] R. Bartsch, E. de Azambuja, I-SPY 2: optimising cancer drug development in the 21st century, ESMO Open. 1 (5) (2016) e000113.
[8] H. Wang, D. Yee, I-SPY 2: a Neoadjuvant adaptive clinical trial designed to improve outcomes in high-risk breast cancer, Curr Breast Cancer Rep. (4) (2019) 303–310, https://doi.org/10.1007/s12609-019-00334-2. Epub 2019 Nov 20. PMID: 33312344; PMCID: PMC7731787.
[9] N. Stallard, L. Hampson, N. Benda, W. Brannath, T. Burnett, T. Friede, P.K. Kimani, F. Koenig, J. Krisam, P. Mozgunov, M. Posch, J. Wason, G. Wassmer, J. Whitehead, S.F. Williamson, S. Zohar, T. Jaki, Efficient adaptive designs for clinical trials of interventions for COVID-19, Statistics in Biopharmaceutical Research 12 (4) (2020) 483–497, https://doi.org/10.1080/19466315.2020.1790415.
[10] RECOVERY Collaborative Group. Azithromycin in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. Lancet. 2021 Feb 13;397(10274):605-612. doi: 10.1016/S0140-6736(21)00149-5. Epub 2021 Feb 2. PMID: 33545096; PMCID: PMC7884931.
[11] J. Dorward, L.M. Yu, G. Hayward, B.R. Saville, O. Gbinigie, O. Van Hecke, E. Ogburn, P.H. Evans, N.P. Thomas, M.G. Patel, D. Richards, N. Berry, M.A. Detry, C. Saunders, M. Fitzgerald, V. Harris, M. Shanyinde, S. de Lusignan, M. I. Andersson, C.C. Butler, Hobbs FR PRINCIPLE Trial Collaborative Group Colchicine for COVID-19 in the community (PRINCIPLE): a randomised, controlled, adaptive platform trial, Br J Gen Pract. 72 (720) (2022) e446–e455, https://doi.org/10.3399/BJGP.2022.0083. PMID: 35440469; PMCID: PMC9037186.
[12] J. Corrigan-Curay, L. Sacks, J. Woodcock, Real-world evidence and real-world data for evaluating drug safety and effectiveness, JAMA 320 (9) (2018) 867–868.
[13] L. Cash-Gibson, J.M. Pericàs, C. Spiertz, E. van de Ketterij, E. Molero, F. Patalano, D. Kalra, A. Ussi, A. Van Dessel, J. Genescà, EU-PEARL: Changing the paradigm of clinical trials in Europe, Eur. J. Pub. Health 31 (Issue Supplement_3) (2021).
[14] REMAP-CAP trial. Available at https://www.remapcap.org/.
[15] HEALY trial. Available at https://clinicaltrials.gov/ct2/show/NCT04297683.
[16] Beat AML trial. Available at https://www.lls.org/bringing-precision-medicine-aml-patients.
[17] M.R. Cowie, J.I. Blomster, L.H. Curtis, S. Duclaux, I. Ford, F. Fritz, S. Goldman, S. Janmohamed, J. Kreuzer, M. Leenay, A. Michel, S. Ong, J.P. Pell, M. R. Southworth, W.G. Stough, M. Thoenes, F. Zannad, A. Zalewski, Electronic health records to facilitate clinical research, Clin Res Cardiol. 106 (1) (2017) 1–9, https://doi.org/10.1007/s00392-016-1025-6. Epub 2016 Aug 24. PMID: 27557678; PMCID: PMC5226988.
[18] The European Council and The European Parliament. 2016. May 4, [2019-02-02]. EUR-Lex. General Data Protection Regulation. Available at https://eur-lex.europa.eu/eli/reg/2016/679/oj.
[19] The Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). Available at https:// https://ohdsi.github.io/CommonDataModel/.
[20] The Observational Health Data Sciences and Informatics program (OHDSI). Available at https://www.ohdsi.org/.
[21] The EHR2EDC project. Available at https://www.i-hd.eu/rd-and-collaborative-projects/ehr2edc/.
[22] B. Harper, N. Christison, Fixing the protocol feasibility process, Available at, Journal of Clinical Research Best Practices (2012) 8, http://clinicalperformancepartners.com/wp-content/uploads/2012/07/Fixing-Feasibility-Final-Jan-2012.pdf.
[23] Pateriya K. Patients Recruitment Forecast in Clinical Trials. August 2015. Available at https://www.cognizant.com/whitepapers/patients-recruitment-forecast-in-clinical-trials-codex1382.pdf.
[24] K.A. Mc Cord, H. Ewald, A. Ladanie, M. Briel, B. Speich, H.C. Bucher, L. G. Hemkens, RCD for RCTs initiative and the Making Randomized Trials More Affordable Group. Current use and costs of electronic health records for clinical trial research: a descriptive study, CMAJ Open 7 (1) (2019) E23–E32, https://doi.org/10.9778/cmajo.20180096. PMID: 30718353; PMCID: PMC6375253.
[25] Y. Ni, J. Wright, J. Perentesis, T. Lingren, L. Deleger, M. Kaiser, I. Kohane, I. Solti, Increasing the efficiency of trial-patient matching: automated clinical trial eligibility pre-screening for pediatric oncology patients, BMC Med Inform Decis Mak. 14 (15) (2015) 28, https://doi.org/10.1186/s12911-015-0149-3. PMID: 25881112; PMCID: PMC4407835.
[26] H.U. Prokosch, T. Ganslandt, Perspectives for medical informatics. Reusing the electronic medical record for clinical research, Methods Inf Med 48 (1) (2009) 38–44.

[27] M. Dugas, M. Lange, C. Muller-Tidow, P. Kirchhof, H.U. Prokosch, Routine data from hospital information systems can support patient recruitment for clinical studies, Clin Trials 7 (2) (2010) 183–189, https://doi.org/10.1177/1740774510363013.
[28] N. Laaksonen, J.M. Varjonen, M. Blomster, A. Palomäki, T. Vasankari, J. Airaksinen, R. Huupponen, M. Scheinin, J. Blomster, Assessing an Electronic Health Record research platform for identification of clinical trial participants, Contemp Clin Trials Commun. 18 (21) (2020 Dec), 100692, https://doi.org/10.1016/j.conctc.2020.100692. PMID: 33409423; PMCID: PMC7773855.
[29] G. De Moor, M. Sundgren, D. Kalra, et al., Using electronic health records for clinical research: The case of the EHR4CR project, J Biomed Inform 53 (2015) 162–173, https://doi.org/10.1016/j.jbi.2014.10.006.
[30] The TriNetX Platform, Available at https://trinetx.com/.
[31] The Clinerion query builder. Available at: https://www.clinerion.com/index/PatientNetworkExplorerSolutions/ClinicalTrialsSolutions/ClinicalApplications/QueryDesigner.html.
[32] ACT Network. Available at: https://www.actnetwork.us/national.
[33] J. Doods, F. Botteri, M. Dugas, F. Fritz, A European inventory of common electronic health record data elements for clinical trial feasibility, Trials 15 (1) (2014) 18, https://doi.org/10.1186/1745-6215-15-18.
[34] I. Reinecke, C. Gulden, M. Kümmel, A. Nassirian, R. Blasini, M. Sedlmayr, Design for a modular clinical trial recruitment support system based on FHIR and OMOP, Stud. Health Technol. Inform. 270 (2020) 158–162, https://doi.org/10.3233/SHTI200142.
[35] 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, M. Sundgren, Federated electronic health records research technology to support clinical trial protocol optimization: Evidence from EHR4CR and the InSite platform, J. Biomed. Inform. 90 (2019), 103090, https://doi.org/10.1016/j.jbi.2018.12.004.
[36] The European Health Data & Evidence Network (EHDEN). Available at https://www.ehden.eu.
[37] G. Hripcsak, J.D. Duke, N.H. Shah, C.G. Reich, V. Huser, M.J. Schuemie, M. A. Suchard, R.W. Park, I.C. Wong, P.R. Rijnbeek, J. van der Lei, N. Pratt, G. N. Norén, Y.C. Li, P.E. Stang, D. Madigan, P.B. Ryan, Observational health data sciences and informatics (OHDSI): Opportunities for observational researchers, Stud. Health Technol. Inform. 216 (2015) 574–578.
[38] M.J. Schuemie, P.B. Ryan, N. Pratt, R. Chen, S.C. You, H.M. Krumholz, D. Madigan, G. Hripcsak, M.A. Suchard, Principles of large-scale evidence generation and evaluation across a network of databases (LEGEND), Journal of the American Medical Informatics Association : JAMIA 27 (8) (2020) 1331–1337, https://doi.org/10.1093/jamia/ocaa103.
[39] I. Li, J. Pan, J. Goldwasser, N. Verma, W.P. Wong, M.Y. Nuzumlalı, B. Rosand, Y. Li, M. Zhang, D. Chang, R.A. Taylor, H.M. Krumholz, D. Radev, Neural natural language processing for unstructured data in electronic health records: A review, Computer Science Review 46 (2022), 100511.
[40] S.M. Meystre, P.M. Heider, Y. Kim, D.B. Aruch, C.D. Britten, Automatic trial eligibility surveillance based on unstructured clinical data, Int J Med Inform. 129 (2019) 13–19, https://doi.org/10.1016/j.ijmedinf.2019.05.018. Epub 2019 May 23. PMID: 31445247; PMCID: PMC6717538.
[41] Investigator Site eSource Readiness Assessment 2020 (eSRA) from eClinicalForum, https://eclinicalforum.org/esra/esra-esource-readiness-assessment-handbook-and-assessment-questionnaire.
[42] TransCelerate eSource Informatics Information Reference Guide for Clinical Research sites 2019, https://www.transceleratebiopharmainc.com/wp-content/uploads/2021/04/TransCelerate_eSource-Informatics-Reference-Guide-for-Clinical-Research-Sites_March-2021.pdf.
[43] the European Medicines Agency Reflection paper on Good Clinical Practice (GCP) compliance in relation to trial master files for management, audit and inspection of clinical trials 2015, https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-good-clinical-practice-compliance-relation-trial-master-files-paper/electronic-management-audit-inspection-clinical-trials_en.pdf.
[44] the Medicines & Healthcare products Regulatory Agency (MHRA) Position Statement and Guidance on Electronic Health Records 2015, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/470228/Electronic_Health_Records_MHRA_Position_Statement.pdf.
[45] J.R. Rogers, G. Hripcsak, Y.K. Cheung, C. Weng, Clinical comparison between trial participants and potentially eligible patients using electronic health record data: A generalizability assessment method, J Biomed Inform. 119 (2021) pp103822, https://doi.org/10.1016/j.jbi.2021.103822. Epub 2021 May 25. PMID: 34044156; PMCID: PMC8260457.
[46] P.S. Brandt, R.C. Kiefer, J.A. Pacheco, P. Adekkanattu, E.T. Sholle, F.S. Ahmad, J. Xu, Z. Xu, J.S. Ancker, F. Wang, Y. Luo, G. Jiang, J. Pathak, L.V. Rasmussen, Toward cross-platform electronic health record-driven phenotyping using Clinical Quality Language, Learn. Health. Syst. 4 (4) (2020) e10233.
[47] R. Farmer, R. Mathur, K. Bhaskaran, S.V. Eastwood, N. Chaturvedi, L. Smeeth, Promises and pitfalls of electronic health record analysis, Diabetologia 61 (6) (2018) 1241–1248, https://doi.org/10.1007/s00125-017-4518-6.
[48] Paulson, M. L., & Weng, C. (2015). Desiderata for Major Eligibility Criteria in Breast Cancer Clinical Trials. In AMIA Annual Symposium Proceedings (Vol. 2015, p. 2025). American Medical Informatics Association.
[49] M. Lehne, J. Sass, A. Essenwanger, et al., Why digital medicine depends on interoperability, Npj Digit Med. 2 (2019) 79, https://doi.org/10.1038/s41746-019-0158-1.
[50] Cohort Diagnostics Package. Available at https://ohdsi.github.io/CohortDiagnostics.
[51] The R Project for Statistical Computing. Available at https://www.r-project.org/.
[52] A.A. Boxwala, B.H. Rocha, S. Maviglia, V. Kashyap, S. Meltzer, J. Kim, R. Tsurikova, A. Wright, M.D. Paterno, A. Fairbanks, B. Middleton, A multi-layered framework for disseminating knowledge for computer-based decision support, Journal of the American Medical Informatics Association : JAMIA 18 (Suppl 1) (2011) i132–i139, https://doi.org/10.1136/amiajnl-2011-000334.
[53] C. Yuan, P.B. Ryan, C. Ta, Y. Guo, Z. Li, J. Hardin, R. Makadia, P. Jin, N. Shang, T. Kang, C. Weng, Criteria2Query: a natural language interface to clinical databases for cohort definition, J Am Med Inform Assoc. 26 (4) (2019) 294–305, https://doi.org/10.1093/jamia/ocy178. PMID: 30753493; PMCID: PMC6402359.
[54] B.N. Wolford, C.J. Willer, I. Surakka, Electronic health records: the next wave of complex disease genetics, Hum. Mol. Genet. 27 (R1) (2018) R14–R21, https://doi.org/10.1093/hmg/ddy081.
[55] H. Ayatollahi, S.F. Hosseini, M. Hemmat, Integrating Genetic Data into Electronic Health Records: Medical Geneticists' Perspectives, Healthcare Informatics Research 25 (4) (2019) 289–296, https://doi.org/10.4258/hir.2019.25.4.289.
[56] K.S. Lau-Min, S.B. Asher, J. Chen, S.M. Domchek, M. Feldman, S. Joffe, J. Landgraf, V. Speare, L.A. Varughese, S. Tuteja, C. VanZandbergen, M.D. Ritchie, K. L. Nathanson, Real-world integration of genomic data into the electronic health record: the PennChart Genomics Initiative, Genetics in Medicine : Official Journal of the American College of Medical Genetics 23 (4) (2021) 603–605, https://doi.org/10.1038/s41436-020-01056-y.