Digital Personalized Health and Medicine
L.B. Pape-Haugaard et al. (Eds.)

doi:10.3233/SHTI200184

Performances of a Solution to Semi-Automatically Fill eCRF with Data from the Electronic Health Record: Protocol for a Prospective Individual Participant Data Meta-Analysis

Nicolas GRIFFON[a,b,1], Helena PEREIRA[c], Juliette DJADI-PRAT[c], María Teresa GARCÍA[d], Sara TESTONI[e], Manon CARIOU[f], Jacques HILBEY[b], Aurèle N'DJA[f] Grégory NAVARRO[g], Nicola GENTILI[e], Oriana NANNI[e], Massimo RAINERI[g], Gilles CHATELLIER[c], Agustín GÓMEZ DE LA CAMARA[d,h], Martine LEWI[g], Mats SUNDGREN[i], Christel DANIEL[a,b], Almenia GARVEY[j], Marija TODOROVIC[g], Nadir AMMOUR[f]

[a] *WIND Department, AP-HP, Paris, France;* [b] *LIMICS, U1142, Inserm, France;* [c] *HEGP, AP-HP, France;* [d] *Hospital 12 de Octubre, Madrid, Spain;* [e] *IRST-IRCCS, Meldola, Italy;* [f] *Sanofi R&D, Paris, France;* [g] *Janssen Pharmaceutica NV, Beerse, Belgium ;* [h] *CIBERESP;* [i] *AstraZeneca, Göteborg, Sweden;* [j] *Icon PLC, Paris, France;*

Abstract. Clinical trial data collection still relies on a manual entry from information available in the medical record. This process introduces delay and error risk. Automating data transfer from Electronic Health Record (EHR) to Electronic Data Capture (EDC) system, under investigators' supervision, would gracefully solve these issues. The present paper describes the design of the evaluation of a technology allowing EHR to act as eSource for clinical trials. As part of the EHR2EDC project, for 6 ongoing clinical trials, running at 3 hospitals, a parallel semi-automated data collection using such technology will be conducted focusing on a limited scope of data (demographic data, local laboratory results, concomitant medication and vital signs). The evaluation protocol consists in an individual participant data prospective meta-analysis comparing regular clinical trial data collection to the semi-automated one. The main outcome is the proportion of data correctly entered. Data quality and associated workload for hospital staff will be compared as secondary outcomes. Results should be available in 2020.

Keywords. Data collection; Health Information Interoperability; Clinical trial as topic; Clinical Trial Protocols as Topic; eSource.

1. Introduction

Pharmaceutical industry is facing increasing costs for drug development [1]. The automatic transfer of data from Electronic Health Records (EHR) to Electronic Data

[1] Corresponding Author, Nicolas Griffon, Web Innovation Données, 33 Bd Picpus – 75012 Paris, France; E-mail: nicolas.griffon@aphp.fr

Capture System (EDC) has been identified as one partial solution to address this issue by reducing the burden of data entry and associated activities, like data monitoring and review by sponsor [2]. Moreover, such solution would drastically reduce the time to entry, which is still too high, even with incentive [3].

A recent literature review [4] identified multiple initiatives towards automated transfer from EHR to EDC systems. Most of these initiatives being monocentric, retrospective or based on only one EHR, the authors also stress out the need for further research to better evaluate the impact of eSource solutions on data quality and workload: some burdens might decrease (entry, quality assessment…) but some others will increase (application maintenance, semantic interoperability maintenance).

The EHR2EDC project [5] has received funding from European Institute of Innovation and Technology (EIT) Health. This project is led by Sanofi R&D, include in a consortium 3 other pharma companies (see Tab 1), one Clinical Research Organization (ICON), one health data technology company (InSite, a TriNetX company), 4 hospitals (see Tab 1), the French National Institute for Medical Research (Inserm) and the European Institute for Innovation through Health Data (i~HD, a not for profit organization). The aim of this project was making such transfer a reality, by overcoming technical, organizational and regulatory difficulties. Building on the already existing InSite solution (developed by InSite), that allows feasibility studies and provides support for recruitment, a new module was developed, that prefills the concomitant medication, local laboratory, vital signs and demographic sections of the eCRF, and sends this information, under the investigator control, to the trial sponsor EDC. After unit, component and usability testing, it is now necessary to evaluate this tool's performances (data transfer capacity, error rate in comparison with manual entry, acceptability with regard to process evolution). This work focusses on the design of the evaluation protocol allowing to assess the performance of an eSource solution and the application of this evaluation protocol in the EHR2EDC context within the TransFAIR study.

2. Methods

2.1. Individual participant data – prospective meta-analysis (IPD–PMA)

IPD meta-analysis is recognized by the Cochrane collaborative group as a 'gold standard' of systematic review [6], and some authors strongly advocate for prospective meta-analysis [7]. Even if this study is a proof-of-concept technology evaluation study, and not a clinical trial, it was decided to design it like an IPD–PMA, following as much as possible PRISMA-P [8][9] and MOOSE [10] statements. This design will ease later inclusion of data from other tests in other hospitals using an eSource solution.

A preliminary protocol was developed by Sanofi R&D and submitted for review and validation to the whole EHR2EDC consortium. Each hospital was allowed to modify the template in order to customize this protocol but agreed to share patient level data to the statistician in charge of the meta-analysis (not the data collected in itself, but rather the result of the comparison of paired data sets collected, either manually or semi-automatically).

2.2. Study design

The main idea of this evaluation is to take advantage of real ongoing clinical trials (support CT), by performing a semi-automatic data collection in addition, and independently, to the usual manual data collection (see Fig 1). Support CT will be conducted in a completely usual way, and will not be affected by the TransFAIR study. The data collected in this way are named "Manual Data". All the patient included in the support CT are eligible for inclusion in the TransFAIR study.

In parallel of each support CT, the EHR2EDC transfer module will be installed allowing direct use of EHR data under the supervision of clinical investigators or study personal in a mirror study. As support CT are ongoing, patients may have been included before the beginning of the TransFAIR study. To increase the amount of data collected, both prospective and retrospective data will be collected using the EHR2EDC transfer module according to the following steps:

1. The study coordinator / investigator will supervise the automated data collection. Through the module interface, (s)he will review, validate and transfer the data required by the protocol to the sponsor eSource database.
2. The sponsor will then reconciliate the manual and eSource database to identify discrepancies and provide them to the study coordinator / investigator.
3. For each discrepancy identified, the study coordinator / investigator will go back to the source documents of both eSource and support CTs data in order to collect the real value of the information collected.
4. To the best of his/her knowledge and skills, (s)he will try to identify error causes amongst: (i) rounding error, (ii) transcription error, (iii) wrong data entry for Manual Data, and (a) modification of information during the transfer, (b) wrong information transferred, (c) user error with the new module (wrong patient, wrong visit or wrong data selection) for the eSource data.

For data outside EHR2EDC scope (e.g. pathology, medical history), Manual Data will only be counted – and considered correct.

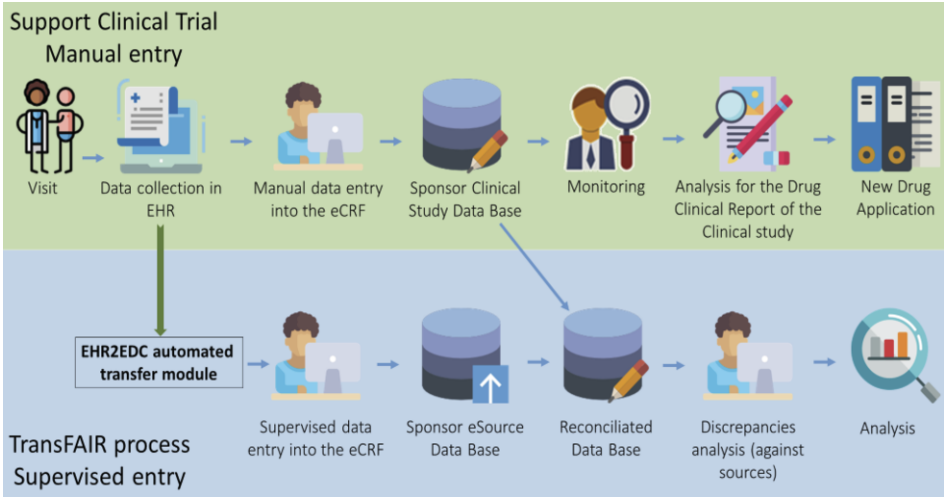

Figure 1. Support CT and TransFAIR study designs.

2.3. Outcomes

The main objective of the TransFAIR study is to assess the percentage of data accurately transferred (the number of data correctly transferred divided by the whole number of correct data – either manually or automatically entered). Secondary outcomes are:

1. Assessment of the percentage of target data points that can be accurately processed by the eSource solution. Computed like main outcome besides the denominator, that is restricted to data transferred.
2. Descriptive analysis of the identified types of inconsistency
3. Comparison of data management activities between the two processes assessed by the number of queries

Subgroup analysis will be performed on study site, type of protocol, medical specialty and data domain. For exploratory analysis, we will:

1. Compare entry error rates between arms, based on the discrepancy analysis, and try to identify some determinants of error rate: datatype (Boolean, text, date…), CDISC domain, data quality…
2. Explore more thoroughly the differences in queries between semi-automated and manual data entry.

2.4. Statistical analysis

We won't assess publication bias because of the prospective design of this meta-analysis. The heterogeneity between sites will be assessed using Cochran's Q and I square statistic. For primary outcome and 1st secondary outcome, proportion will be computed using a multi-level model to take account of inter site variance. A conditional logistic regression model will be used to compute OR estimate for entry error between arms. Once again, the inter site heterogeneity will be handled by a multi-level model.

2.5. Ethical concerns

Patients included in the support clinical trials will be asked if they consent to the reuse of their data in the context of the TransFAIR study. All the data shared between partners are shared with respect to the EU General Data Protection Regulation (GDPR).

3. Results

The TransFAIR study is being conducted using the proposed study protocol at 3 hospitals in parallel of 6 support clinical trials between mid-September and end-November, 2019 (see Tab 1). The 3 IRBs of the hospitals involved in the project approved both TransFAIR and support CT.

4. Discussion

Making the data interoperable between pharma and hospital world is a huge work that has been accomplished for demographic, local laboratory, medication and vital signs data

during the EHR2EDC project. It required the creation – to be published – of: (i) a FHIR common information model (CIM) – each hospital must be able to provide data in the CIM way, and, (ii) an important set of mappings between this CIM and pharma standards – derived from CDISC. This model makes different EHR and EDC systems interoperable.

Table 1. Description of support clinical trials

Running hospitals	Registration number[1]	Sponsor	Indication	Phase
APHP[2], 12Oct	NCT03767244	Janssen	Advanced prostate cancer	3
APHP, IRST-IRCCS	NCT03390504	Janssen	Advanced urothelial cancer	3
APHP	NCT03315143	Sanofi	Diabetes/cardiology	3
12Oct	NCT03284957	Sanofi	Advanced breast cancer	1b
12Oct	NCT03619213	Astra Zeneca	Heart failure	3b
IRST-IRCCS	NCT02516241	Astra Zeneca	Urothelial cancer	3

1: ClinicalTrials.gov identifier; 2: four APHP hospitals involved (Mondor, HEGP, Bichat, Lariboisière)

This study will allow the identification of data quality gaps in EHR, highlighting the data quality dimensions, as described by Khan et al., to put the focus on for the EHR2EDC scenario. The idea is to enhance the i~HD quality seal for Research Platforms [12] and to make it a requirement for any hospital to deploy and use any data transfer application between EHR and EDC.

This paper focused on the design of an evaluation protocol allowing the assessment of eSource solutions implementing large-scale connectivity between EHR and EDC systems. Results will be available in 2020.

References

[1] DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs. J Health Economics 47 (2016), 20-33.

[2] Eisenstein EL, Nordo AH, Zozus MN. Using Medical Informatics to Improve Clinical Trial Operations. Stud Health Technol Inform. 234 (2017), 93-97.

[3] Ahonen K, Herrezeel T, Hull J, King R, Nadolny P and Perkins R. 2001 to 2009: Investigational Site Perspectives on Clinical Trial Information Systems. eClinical Forum 2012. www.eclinicalforum.org

[4] Garza M, Myneni S, Nordo A, Eisenstein EL, Hammond WE, Walden A, et al. eSource for Standardized Health Information Exchange in Clinical Research: A Systematic Review. Stud Health Technol Inform. 257 (2019), 115 124

[5] EHR2EDC. Available at: http://www.eithealth.eu/ehr2edc [last check 30/09/2019]

[6] About IPD meta-analyses. Available at https://methods.cochrane.org/ipdma/about-ipd-meta-analyses [last check 30/09/2019]

[7] Ioannidis J. Next-generation systematic reviews: prospective meta-analysis, individual-level data, networks and umbrella reviews. Br J Sports Med. 51(20) (2017), 1456–1458.

[8] Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols(PRISMA-P) 2015 statement. Syst Rev.4 (2015), 1.

[9] Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols(PRISMA-P) 2015: elaboration and explanation. BMJ. 349 (2015), g7647.

[10] Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al for the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. Meta-analysis of observational studies in epidemiology: a proposal for reporting. JAMA 283 (2000), 2008–12.

[11] Kahn, M. G., Callahan, T. J., Barnard, J., Bauck, A. E., Brown, J., Davidson, B. N., ... & Liaw, S. T. (2016). A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. Egems, 4(1).

[12] The i~HD Quality Seal for Research Platforms (QS4RP). Available at: https://www.i-hd.eu/index.cfm/services/qs4rp/ [last check 30/09/2019]