Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240679

HCE2RNFC: An Efficient Methodology for Reusing the EHR in the Spanish National Hip Fracture Registry

Paula RUBIO-MAYO [a, b, 1], Cristina OJEDA-THIES [a, c], María Jesús JIMÉNEZ-CEREZO [a], Noelia GARCIA-BARRIO [a], Juan Luis CRUZ-BERMÚDEZ [a] and Miguel PEDRERA-JIMÉNEZ [a, b, d]
[a] *Digital Health Group, Research Institute Hospital 12 de Octubre, Madrid, Spain*
[b] *ETSI Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain*
[c] *Department of Traumatology and Orthopedic Surgery, Hospital Universitario 12 de Octubre, Madrid, Spain*
[d] *Better, Ljubljana, Slovenia*
ORCiD ID: Paula Rubio-Mayo https://orcid.org/0000-0001-5114-4344

Abstract. Hip fracture is a condition associated with ageing and frailty, with an associated prevalence of 7 per 10000 population in Spain. Evidence suggests that factors in the healthcare process can influence clinical outcomes, so the creation of a national registry is an opportunity to monitor and improve this process. In this regards, Electronic Health Record (EHR) can provide a large amount of data, that can be used to populate the Spanish National Hip Fracture Registry (RNFC, by its acronym in Spanish). However, this reuse of the EHR requires a prior effort in modelling and standardization to build the extraction, transformation, and loading (ETL) processes in a flexible, transparent, and scalable manner. In this work, a robust EHR reuse methodology is implemented to obtain EHR-derived data for the RNFC. The main result of this work was the design and implementation of an EHR data reuse methodology, which was able to load 1279 hip fracture cases and almost 68% of the required concepts from the RNFC.

Keywords. Electronic Health Record, Hip Fracture, National Registry, Real World Data, Semantics, Standards, Health Data Space.

1. Introduction

Hip fracture is a condition associated with ageing and frailty [1,2] and there is evidence that factors in the healthcare process influence its development and clinical outcomes [2-4]. The Spanish National Hip Fracture Registry (RNFC, by its acronym in Spanish) is a rich and harmonized data source for generating statistics and metrics for the Spanish National Health System [2], which is manually loaded by each clinician involved in the different nodes of the network. This national registry provides an opportunity to perform

[1] Corresponding Author: Paula Rubio Mayo, Hospital Universitario 12 de Octubre, Av. De Córdoba s/n, 28041, Madrid Spain; E-mail: paula.rubio.externo@salud.madrid.org.

clinical audit by assessing weaknesses, identifying opportunities for improvement, and monitoring their impact on clinical and management outcomes [2].

In this scenario, the Electronic Health Record (EHR) can provide a huge amount of data to be reused in care planning and research [5-7]. Thus, these EHR-derived data allow to assess the real impact of the treatments and diagnostics in different real-world settings [7,8]. However, the heterogeneity of data collected processes for each institution combined with the primary purpose of the collection process, requires a previous effort for building extraction, transformation, and loading (ETL) methodology in a flexible, transparent, and scalable manner [5,6,9].

Therefore, this work aims to design and implement a methodology for reusing the EHR in the RNFC, ensuring efficiency and data quality.

2. Methods

The methodology developed is divided into three stages and its workflow is described in Figure 1. First, labelling of the hip fracture cohort using standard terminology for subsequent identification. Second, the identification of the essential clinical concepts of the RNFC available in the EHR in a structured and standardized way. And third, the identified data elements are transformed into the format required by the national registry and loaded into its database.

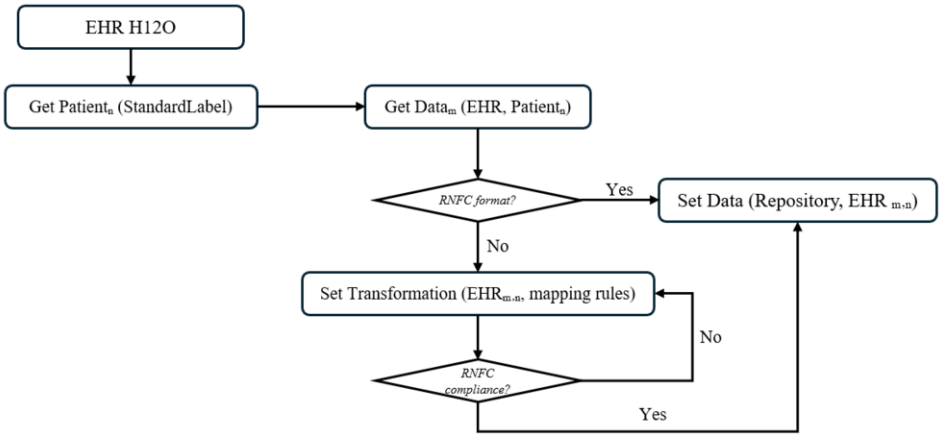

Figure 1. Algorithm for the reuse of data from the electronic health record for the completion of the hip fracture registry (RNFC), Abbreviation: H12O: Hospital Universitario 12 de Octubre.

The data transformations were applied following the previously defined methodology based on the definition of transparent, standardized, and reusable operations [5] and were validated and verified with physicians in order to ensure end-user understanding.

2.1. Identification of the Hip Fracture cohort

Identification of the cohort in this study was possible due to the standardized labeling of the cohort [10]. The terminological concept from SNOMED CT [11] associated with the diagnosis of hip fracture is "263225007 – Proximal femur fracture (disorder)", which was implemented in the EHR at 2022-04-30. The clinician associated manually the code to the acute care episode upon diagnosing a case of hip fracture. This allowed to extract, in real time, all the data associated with the cohort and build the ETL process in the RNFC repository.

2.2. Modeling and standardization of the EHR

The data extraction source was the InfoBanco platform of H12O [12], which centralized all EHR data of the organization. In this platform, the health information is modelled and persisted according to OpenEHR [13], using Better Platform [14], and incorporates terminologies such as SNOMED CT [11] and LOINC [15]. This standardization of the EHR has allowed their reuse, without additional manual efforts and full meaning, in data collection processes for research and other secondary uses. All the concepts required by the harmonized model of the RNFC were identified in the InfoBanco platform and subsequently mapped to the RNFC format. In Table 1 describes the main tables containing essential information for extracting the required concepts and their corresponding terminology.

Table 1. Information models for EHRs reuse in InfoBanco Platform.

Archetype	Description	Terminology binding
Patient	Demographics data, e.g., birthdate, sex and vital status.	SNOMED CT
Encounter	Data related to inpatient, emergency and outpatient visits.	SNOMED CT
Location	Patient locations during hospitalization, e.g., ICU admission.	SNOMED CT
Observation	Clinical, laboratory and patient-reported observations.	SNOMED CT, LOINC
Diagnosis	Health issues and clinical diagnoses.	SNOMED CT
Medication	Pharmacological treatment prescribed.	SNOMED CT
Procedure	Procedures performed, e.g., surgeries and nursing interventions.	SNOMED CT

2.3. ETL process to the RNFC repository

Firstly, extraction and transformation stages were developed in RStudio. On the one hand, the extraction process involved a set of queries launched on the main tables of the InfoBanco platform [12]. On the other hand, the transformation process consisted of applying a set of operations previously defined to the concepts of the EHR to obtain the required concepts in the RNFC format. Finally, obtained data was loaded into REDCap, a secure web application for building and managing online surveys and databases [16]. Therefore, for this work, a REDCap database with forms of the RNFC was configured to be automatically populated with data from the EHR.

3. Results

The main result of this work was the design and implementation of a reliable methodology for EHR data reuse that has enabled the loading of 1279 hip fracture

condition episodes from 1253 patients between August 2021 and February 2024. This reuse strategy allows the automation of the extraction, transformation, and loading (ETL) processes. This reduces the time required by the clinicians in the data collection process, in the hip fracture scenario.

Table 2 shows that a large volume of records and concepts were obtained for the identified cohort, using this methodology, for each form of the RNFC. Nearly 68% of the concepts are already available in structures form in the EHR and can be reused.

Table 2. Completeness analysis result.

RNFC section	Records (N)	Patients (N)	Concepts (%)
Demographic	6516	1253	100
Characteristics of patients	7042	1253	50
Acute handling	7701	1253	60
Discharge	8646	1253	90
30-day follow-up	6516	1253	63.63
120-day follow-up	6516	1253	63.63

Likewise, the times of data collection compared to traditional manual collection were analyzed. It could be estimated that the time necessary to verify the data obtained automatically was 6-7 minutes per clinical case, while the time reported by professionals to enter the data manually was 13-14 minutes per clinical case.

4. Discussion

The benefits of reusing data from the EHR are many, most notably reducing the time clinicians spend collecting data by almost half. Simultaneously, this reduction in time translates into a reduction in human resources and economic costs. Nevertheless, it is a goal to ensure that such methodologies are transparent, understandable, and trustworthy. Only by achieving that these ETL processes are understandable, will the feasibility of further studies with reused data be assured [6].

Similarly, the new digital era requires a change of the mindset to make the EHR less reliant on unstructured data, and to follow terminologies and standards that will allow it to be interoperable between institutions and facilitate the work of identifying key concepts and creating cohorts [5, 6, 10]. By standardizing processes and improving the transparency of data management, collaborative research, and generation of real-word evidence will be encouraged, helping to achieve significant advances that could be considered milestones in clinical and medical research [9].

5. Conclusions

The standardization of the hip fracture condition enabled the identification of 1279 episodes, associated with 1253 patients from August 2021 to February 2024. The reuse of data in this work allowed loading almost 68% of the required concepts of the RNFC for 1279 hip fracture condition episodes, with a significant reduction in the time required for researchers to complete the forms. Thus, this study could demonstrate the effectiveness of this ETL methodology in a hip fracture scenario, facilitating the participation of the H12O in a national registry in less burdensome way.

Future work includes a more complete analysis of the data quality dimensions where the consistency and completeness of the reused data could be objectively assessed, and a time analysis including development code time. Finally, a comparative study will be conducted to evaluate the impact of this methodology against a manual repository.

References

[1] Hip fracture healthcare in the hospitals of the Spanish National Health System. Available from: https://www.sanidad.gob.es/estadEstudios/estadisticas/docs/Estadisticas_comentadas_01.pdf Accessed 2024 May 24

[2] Technical Manual of the Spanish National Hip Fracture Registry. Available from: https://rnfc.es/ Accessed 2024 May 25

[3] Liem IS, Kammerlander C, Suhm N, et al. Identifying a standard set of outcome parameters for the evaluation of orthogeriatric co-management for hip fractures. Injury. 2013;44(11):1403-1412. doi:10.1016/j.injury.2013.06.018

[4] Alarcón Alarcón T, González-Montalvo JI. Fractura osteoporótica de cadera. Factores predictivos de recuperación funcional a corto y largo plazo [Osteoporotic hip fracture. Predictive factors of short-and long-term functional recovery]. An Med Interna. 2004;21(2):87-96. doi:10.4321/s0212-71992004000200010

[5] Pedrera-Jiménez M, García-Barrio N, Rubio-Mayo P, et al. TransformEHRs: a flexible methodology for building transparent ETL processes for EHR reuse. Methods Inf Med. 2022;61(S 02):e89-e102. doi:10.1055/s-0042-1757763

[6] Pedrera M, Garcia N, Rubio P, Cruz JL, Bernal JL, Serrano P. Making EHRs Reusable: A Common Framework of Data Operations. Stud Health Technol Inform. 2021;287:129-133. doi:10.3233/SHTI210831

[7] Declerck J, Kalra D, Vander Stichele R, Coorevits P. Frameworks, Dimensions, Definitions of Aspects, and Assessment Methods for the Appraisal of Quality of Health Data for Secondary Use: Comprehensive Overview of Reviews. JMIR Med Inform. 2024;12:e51560. Published 2024 Mar 6. doi:10.2196/51560

[8] Kahn MG, Callahan TJ, Barnard J, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. EGEMS (Wash DC). 2016;4(1):1244. Published 2016 Sep 11. doi:10.13063/2327-9214.1244

[9] EHR reuse methodology applied to the national hip fracture registry. Available from: https://www.researchgate.net/publication/380876838_METODOLOGIA_DE_REUTILIZACION_DE_LA_HCE_APLICADA_AL_REGISTRO_NACIONAL_DE_FRACTURA_DE_CADERA_RNFC Accessed 2024 May 25

[10] Pedrera-Jiménez M, García-Barrio N, Cruz-Rojo J, et al. Obtaining EHR-derived datasets for COVID-19 research within a short time: a flexible methodology based on Detailed Clinical Models. J Biomed Inform. 2021;115:103697. doi:10.1016/j.jbi.2021.103697

[11] SNOMED International. Available from: https://www.snomed.org/ Accessed 2024 May 25

[12] Infobanco Platform. Available from: https://www.nttdata.com/global/en/insights/focus/the-infobanco-health-data-platform-from-innovative-idea-to-consumer-success-within-one-year Accessed 2024 May 25

[13] Pedrera-Jiménez M, García-Barrio N, Frid S, et al. Can OpenEHR, ISO 13606, and HL7 FHIR Work Together? An Agnostic Approach for the Selection and Application of Electronic Health Record Standards to the Next-Generation Health Data Spaces. J Med Internet Res. 2023;25:e48702. Published 2023 Dec 28. doi:10.2196/48702

[14] Better Platform. Available from: https://www.better.care/better-platform/ Accessed 2024 May 25

[15] LOINC. Available from: https://loinc.org/ Accessed 2024 May 25

[16] REDCap. Available from: https://www.project-redcap.org/ Accessed 2024 May 25