Automated Generation of Individual and Population Clinical Pathways with the OMOP Common Data Model

Fabio BOUDIS[a,b,1], Guillaume CLEMENT[a], Amelie BRUANDET[a,c] and Antoine LAMER[a,b,c]

[a] *CHU Lille, Service Analyse et Traitement de l'information médicale, F-59000, Lille, France*
[b] *Univ. Lille, Faculté Ingénierie et Management de la Santé, F-59000, Lille, France*
[c] *Univ. Lille, CHU Lille, ULR 2694 - METRICS: Évaluation des Technologies de santé et des Pratiques médicales, F-59000, Lille, France*

Abstract. *Introduction*. Clinical pathways represents the sequence of interventions from which the patients benefit during their encounters with health care structures. There are several complex issues which make it difficult to represent these pathways (e.g. high numbers of patients, heterogeneity of variables). *Methods*. We developed a tool to automate the representation of clinical pathways, from an individual and population points of view, and based on the OMOP CDM. The tool implemented the Sankey diagram in three stages: (i) data extraction, (ii) generation of individual sequence of steps and (iii) aggregation of sequence to obtain the population-level diagram. We tested the tool with three surgery procedures : the total hip replacement, the coronary bypass and the transcatheter aortic valve implantation. *Results*. The tool provided different ways of visualizing pathways depending on the question asked: a pathway before a surgery, the pathway of deceased patients or the complete pathway with different steps of interest. *Discussion*. We proposed a tool automating the representation of the clinical pathways, and reducing complexity of visualization. Representations are detailed from an individual and population points of view. It has been tested with three surgical procedures. The tool functionalities will be extended to cover a greater number of use cases.

Keywords. Pathway, Data visualization, Data reuse, OMOP, Sankey diagram

1. Introduction

Clinical pathways represents the sequence of interventions from which the patients benefit during their encounters with health care structures [1,2]. They are used in two contexts: an individual-based context when the visualization targets a single patient to provide appropriate care, and a population-based context when it relates to a whole population and addresses research or decision-making objectives [2]. There are several complex issues which make it difficult to represent these pathways. First, the high numbers of patients and the heterogeneity of variables increase the complexity of datasets and induce information overload. Second, pathway data contains temporal

---

[1] Corresponding Author, Fabio Boudis, Service Archives et Traitement de l'information médicale, CHU Lille, 2 avenue Oscar Lambret, F-59000, Lille, France; E-mail: fabio.boudis.etu@univ-lille.fr.

events that need to be aligned for population-based data analysis. Existing solutions have some limitations. They were often developed for representing the pathway of a selected pathology or a single patient, or they used data stored in a handmade data model, which complicates the use of the tool in other contexts or with other database formats [2-6].

Observational Health Data Sciences and Informatics (OHDSI) developed the Observational Medical Outcomes Partnership (OMOP) common data model (CDM). It allows to standardize the data structure and vocabulary of data, which are originally collected by different software and characterized by heterogeneous vocabularies, depending on the country [7]. Based on the CDM, it is possible to share tools, methods and results. The OHDSI consortium thus shares open access R packages, ETL tools and scripts for performing reproducible statistical analyses and evidence-based research [8].

The objective of this work is to propose a tool to automate the representation of the clinical pathways, from an individual and population points of view, based on the OMOP CDM.

2. Methods

2.1 Concepts

In a previous study, we identified and used the Sankey diagram to represent a sequence of steps [9]. The Sankey diagram is a flow diagram composed of nodes and links between nodes. In the hospital field, the nodes correspond to steps or events and are labeled according to unit care, drugs administrations, medical procedures, diagnoses. The links represent a set of patients going from one step to another. The width of nodes and links is proportional to the number of patients. Generally, a step that all patients go through is used to align the diagram.

2.2 Implementation

Data were extracted from the database of the Lille University Hospital and covering hospital stays, medical procedures and diagnoses, for year 2019. Local data was transformed into the OMOP CDM before processing.

In a first step, concepts corresponding to inclusion criteria were selected to filter the patients. Data were extracted from visit_detail, condition_occurrence, procedure_occurrence and concept_relationship tables.

The second step dealt with the generation of the individual pathways. For this, the temporal sequence of steps is reproduced, in integrating the inclusion event. Four lists of indices are developed : in starting from the first step of the sequence (list A), in starting from the last step of the sequence (list B), in starting from the inclusion event by moving to the endpoints (list C) and in starting from the endpoints to the inclusion event (list D). The steps of interest are selected according to their indices in the sequence of steps. An example of the lists of indices is presented in Figure 1. The steps labels may be simplified with mapping from concept_relationship table (e.g. all intensive care units are mapped to a unique intensive care unit label).

In the last step, individual pathways are gathered to obtain population pathways. The size of the nodes represents the population in the step. The flows between the nodes represent the size of the population moving from one step to another. The Sankey

diagram is indexed on the inclusion event, and only previously selected steps are represented.

The tool was developed with R (version 4.0.3) and uses DatabaseConnector, SqlRender, d3js, plotly, tidyverse, and lubridate packages. Our R package for the generation of clinical pathways is available here : https://gitlab.com/Fboudis/patient_pathway.

List A	1	2	3	4	5	6
List B	-6	-5	-4	-3	-2	-1
Sequence	Step 1	Step 2	Inclusion Event	Step 3	Step 4	Step 5
List C	-2	-1	0	1	2	3
List D	1	2	0	-3	-2	-1

Figure 1. Sequence of steps and inclusion event, and the related indices lists to identify a step.

2.3 Study cases

We choose 3 study cases in connection with the issues we routinely handle in our hospital : the total hip surgery (SC1), the coronary bypass (SC2) and the transcatheter aortic valve implantation (SC3). These 3 procedures benefit from different care pathways, and we wished to visualize the steps prior to surgery in SC1, to study the pathway of dead patients in SC2 and to visualize the whole hospital stay in SC3.

3. Results

We tested our package with 1013, 566, and 357 patients for SC1, SC2, and SC3 respectively. Due to label simplification, the number of distinct labels have been reduced from 31 to 11, from 37 to 12, and from 43 to 10 for SCI, SC2 and SC3 respectively. Due to simplification, the number of unique individual pathways have been decreased from 43 to 41, from 188 to 145, from 179 to 124 for SCI, SC2 and SC3 respectively. The Figure 2 represents the 3 Sankey diagrams for SC1 (A), SC2 (B), and SC3 (C).

In SC1, there were two main flows before surgery: (i) patients came from home and were admitted in conventional or weekly units, or (ii) patients were admitted in emergency units for unplanned admission.

In SC2, all patients went directly to intensive care units. The diagram showed that some of the deceased patients had returned to continuous care before their situation deteriorated and died, or got readmitted to intensive care before dying (orange flow, Figure 2 - B).

In SC3, the diagram describes the global pathway before and after the surgery. Patients went in several types of units, in equivalent proportions. It highlighted the heterogeneity of the care. Around a quarter of patients were transferred to another hospital at the end of the stay.
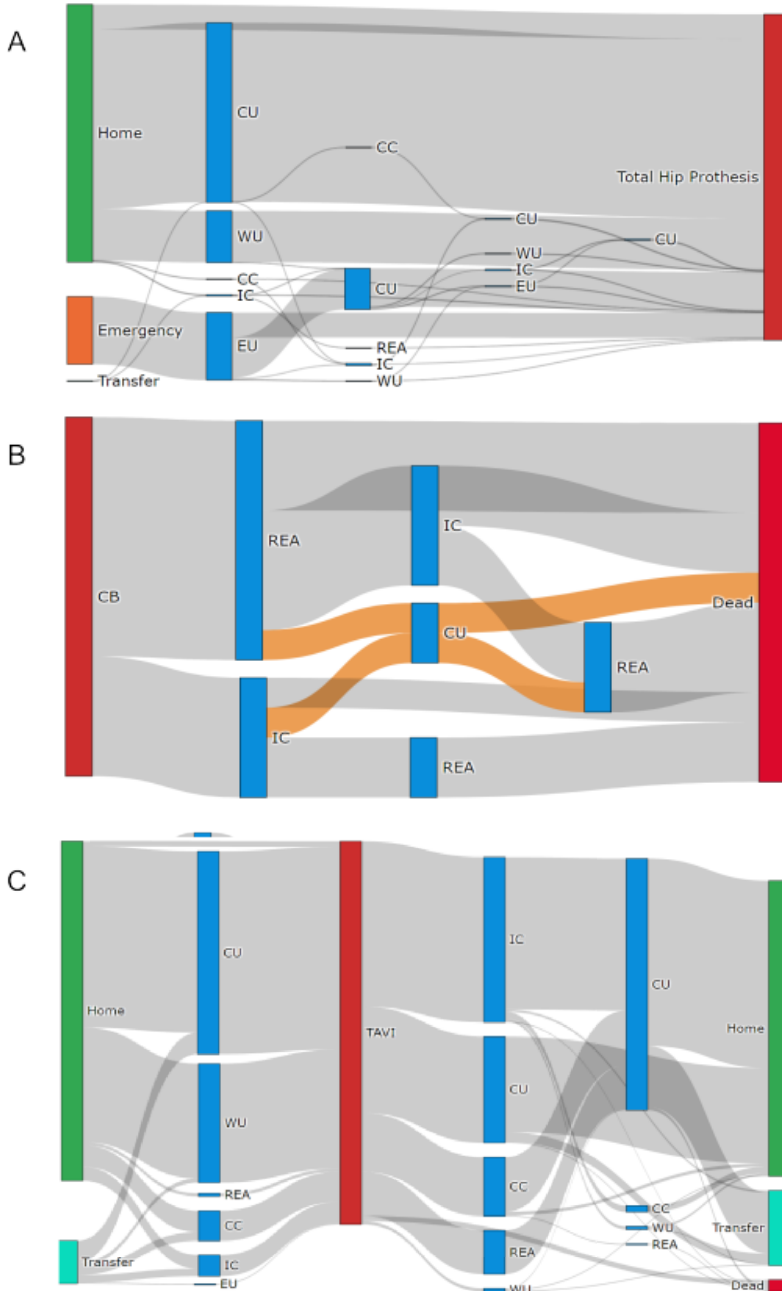
Figure 2. Sankey diagrams for : A – Total hip prosthesis, B – Coronary bypass, C - Transcatheter aortic valve implantation. CB: Coronary bypass; CC: Continuous care; CU: Conventional unit; EU: Emergency unit; IC: Intensive care; REA: Reanimation; WU: Weekly unit.

4. Discussion

In this paper, we described a tool to automate the representation of the clinical pathway. This tool is in open access on our git repository and supports databases in the OMOP CDM. By proposing the selection of steps through different index lists, and by simplifying the heterogeneous labels, our tool helps to reduce the complexity of the patient's journey. The tool has been tested with three different surgical operations, each time focusing on a different part of the hospital stay. The graphical representation at patient level allows the clinician to have an overview of the hospital stay of a selected patient. The representation at the population level offers the possibility to identify atypical care pathways, forecasting the resources to be allocated, and conducting research. In this study, only inpatient stays were considered and we do not take into account outpatient study cases. The index date was defined according to a procedure, and could take into account other events such as drug administration or diagnosis. Further evaluation of the tool is needed, in particular with clinicians and researchers to ensure the contribution of sequence simplification methods.

5. Conclusions

We proposed a tool to automate the representation of the clinical pathways, from an individual and population points of view, based on the OMOP CDM. It has been tested with three surgical procedures. The tool functionalities will be extended to cover a greater number of use cases.

References

[1]   Perer A, Wang F, Hu J. Mining and exploring care pathways from electronic medical records with visual analytics. J Biomed Inform. 2015 Aug;56:369-78.

[2]   Martignene N, Balcaen T, Bouzille G, Calafiore M, Beuscart JB, Lamer A, Legrand B, Ficheur G, Chazard E. Heimdall, a Computer Program for Electronic Health Records Data Visualization. Stud Health Technol Inform. 2020 Jun 16;270:247-251. doi: 10.3233/SHTI200160. PMID: 32570384.

[3]   Goren-Bar D, Shahar Y, Galperin-Aizenberg M, Boaz D, Tahan G, KNAVE II: The Definition and Implementation of an Intelligent Tool for Visualization and Exploration of Time-oriented Clinical Data, in: Proc. Work. Conf. Adv. Vis. Interfaces, ACM, New York, NY, USA; 2004. p. 171–174.

[4]   Plaisant C, Mushlin R, Snyder A, Li J, Heller D, Shneiderman B, LifeLines: Using Visualization to Enhance Navigation and Analysis of Patient Records, in: B.B. Bederson, and B. Shneiderman (Eds.), Craft Inf. Vis., Morgan Kaufmann, San Francisco; 2003. p. 308–312.

[5]   Bui AAT, Aberle DR, Kangarloo H, TimeLine: Visualizing Integrated Patient Records, IEEE Trans. Inf. Technol. Biomed. 2007;11:462–473.

[6]   Wongsuphasawat K, Gotz D, Exploring Flow, Factors, and Outcomes of Temporal Event Sequences with the Outflow Visualization, IEEE Trans. Vis. Comput. Graph. 2012;18:2659–2668.

[7]   Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, Suchard MA, Park RW, Wong IC, Rijnbeek PR, van der Lei J, Pratt N, Norén GN, Li YC, Stang PE, Madigan D, Ryan PB. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. Stud Health Technol Inform. 2015;216:574-8.

[8]   Reps JM, Schuemie MJ, Suchard MA, Ryan PB, Rijnbeek PR. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. J Am Med Inform Assoc. 2018;25(8):969-975.

[9]   Lamer A, Laurent G, Pelayo S, El Amrani M, Chazard E, Marcilly R. Exploring Patient Path Through Sankey Diagram: A Proof of Concept. Stud Health Technol Inform. 2020 Jun 16;270:218-222. doi: 10.3233/SHTI200154. PMID: 32570378.