RESEARCH

Open Access

Feasibility of automated surveillance of implantable devices in orthopaedics via clinical data warehouse: the Studio study

Marie Ansoborlo[1,2,6], Christine Salpétrier[2,6], Louis-Romé Le Nail[4,5], Julien Herbet[2,6], Marc Cuggia[3], Philippe Rosset[4,5] and Leslie Grammatico-Guillon[1,2,6*]

Abstract

Background Total hip, knee and shoulder arthroplasties (THKSA) are increasing due to expanding demands in ageing population. Material surveillance is important to prevent severe complications involving implantable medical devices (IMD) by taking appropriate preventive measures. Automating the analysis of patient and IMD features could benefit physicians and public health policies, allowing early issue detection and decision support. The study aimed to demonstrate the feasibility of automated cohorting of patients with a first arthroplasty in two hospital data warehouses (HDW) in France.

Methods The study included adult patients with an arthroplasty between 2010 and 2019 identified by 2 data sources: hospital discharge and pharmacy. Selection was based on the health insurance thesaurus of IMDs in the pharmacy database: 1,523 distinct IMD references for primary THSKA. In the hospital discharge database, 22 distinct procedures for native joint replacement allowing a matching between IMD and surgical procedure of each patient selected. A program to automate information extraction was implemented in the 1st hospital data warehouse using natural language processing (NLP) on pharmacy labels, then it was then applied to the 2nd hospital.

Results The e-cohort was built with a first arthroplasty for THKSA performed in 7,587 patients with a mean age of 67.4 years, and a sex ratio of 0.75. The cohort involved 4,113 hip, 2,630 knee and 844 shoulder surgical patients. Obesity, cardio-vascular diseases and hypertension were the most frequent medical conditions.

Discussion The implementation of an e-cohort for material surveillance will be easily workable over HDWs France wild. Using NLP as no international IMD mapping exists to study IMD, our approach aims to close the gap between conventional epidemiological cohorting tools and bigdata approach.

Conclusion This pilot study demonstrated the feasibility of an e-cohort of orthopaedic devices using clinical data warehouses. The IMD and patient features could be studied with intra-hospital follow-up and will help analysing the infectious and unsealing complications.

Keywords Medical device safety, Epidemiologic surveillance, Data warehousing, Arthroplasty

*Correspondence:
Leslie Grammatico-Guillon
leslie.guillon@univ-tours.fr
[1]Health Ethics EA, Medecine Faculty, Tours University, Tours
F-37000, France
[2]Clinical Data Centre, CHRU Tours, Tours F-37000, France
[3]Univ Rennes, CHU Rennes, Inserm, LTSI – UMR 1099, Rennes
F-35000, France
[4]Orthopaedics Surgery and Traumatology Departement, CHRU de Tours,
Tours F-37000, France
[5]HUGORTHO network, Tours, France
[6]Department of Public Health, University hospital of Tours, Tours, France

Background

Population ageing is associated with increasing orthopaedic surgery due to disability and functional disorders leading to an expansion of demands [1]. in France more than 200,000 total arthroplasties for hip (THA), knee (TKA) and shoulder (TSA) are performed annually [2, 3]. In this context, patient conditions (age, polymedication, comorbidities) raise the risk of complications such as infections (1% for THA and TKA) and thrombosis (2 to 3% for THA and 9 to 12% for TKA) [4–6]. Even though serious side effects are rare [7], THKSA (total hip, knee or shoulder arthroplasty) could be responsible for longer stays, poorer quality of life, incapacity, rehospitalisation and second surgery, hence representing a public health issue [8–10].

Although patient characteristics associated with complication are well-established [11–14], accurately assessing the impact of implantable medical devices (IMDs) or surgical techniques on complication occurence remains challenging. It has been sparsely explored, notably because of the rarity of such events, requiring large study populations. Orthopaedic surveillance repositories have been implemented since 1975, but they are often limited by high costs and incomplete data, particularly in France [13, 14]. Their goals are not fully met, partly due to insufficient information regarding data quality and coverage [15, 16]. Thus, larger cohort studies are needed to address physician concerns and regulatory authority requirements and to guide public health decision-making [15]. This is particularly relevant given the low frequency of implant-related issues and the recent updated requirements in material-vigilance [17, 18].

In Western France, interoperable Clinical Data warehouses (CDW) have been implemented in 6 hospitals, offering potential for improving device surveillance by the reuse of real-life clinical data [19, 20]. These CDWs integrate multiple sources: (i) hospital discharge database from the French Diagnosis Related Groups (DRG) which encompasses the international classification of diseases (ICD) and the French Current Procedural Terminology (CPT), (ii) Electronic Health Records (prescriptions for bio-clinical examinations and results, consultation and hospitalisation reports, etc.) and (iii) pharmacy records (medication, devices, etc.). Data collection automation through CDW real-world digital data could represent a great opportunity to improve medical device monitoring. A recent literature review on the use of CDWs for IMD surveillance highlighted several ongoing challenges, such as data quality and the need for improved semantic interoperability, requiring the adoption of standardized, regularly updated reference terminologies. Additionally, the review underscored the importance of inserting a Unique Device Identifier (UDI) and detailed annotations on IMDs to enhance surveillance effectiveness [21].

Currently, there is still no sufficiently complete or qualitative IMD knowledge database to meet this need, but efforts are ongoing. However, real-world digital data offer many opportunities for improving the automation of monitoring in orthopaedics: structured data could be used to describe both IMD and patient. The contribution of real-world data, especially through Natural language processing (NLP) and the integration of device Knowledge databases intended to better describe IMDs could improve current surveillance performance (17–19).

This study, so called STUDIO (*STUdy of automate surveillance of Devices In Orthopedics via big data tools*), aimed to assess the feasibility of creating an e-cohort of patients undergoing THKSA using French CDWs. The ultimate goal is to provide surgeons and regulatory authorities with insights into material safety by leveraging real-life clinical data.

Methods

The STUDIO study was an exploratory pilot designed to assess the feasibility of developing and utilizing a THKSA e-cohort. First, the cohort population was identified from two hospital knowledge databases. The patient characteristics were then described and finally, IMDs were characterised with a focus on THA devices.

Patient inclusion in the e-cohort

The population concerned all the hospital stays of adult patients for a primary THKSA between 2010 and 2019 in the orthopaedic departments of two university hospitals in Western France (centre I and centre II) having implemented a CDW based on the eHOP® model [19]. To accurately identify the cases, two data sources were crossed, data from the (i) hospital pharmacy database and (ii) hospital discharge database. Hospital stays were selected only if data existed and matched from the two sources (i.e. same joint). Patients could possibly have multiple stays for a joint replacement and each one was included (e.g.: a patient with one stay for THA and one for TKA or with one stay for a left hip procedure and one for the right hip). To focus on the joint replacement procedures, hospital stays with a primary diagnosis of fracture, cancer or infection were excluded. Patients excluded from the e-cohort were described by exclusion reason (fracture, cancer or infection), by joint (THKSA) and by care centre (I and II) (supplementary data, figure A1).

Hospital pharmacy database

The hospital pharmacy database records the LPP code for each device. This mandatory information is useful for IMD companies, national institutions and healthcare facility pharmacies to identify every component of an orthopaedic device; they are linked to the French Health Insurance system, according to French regulatory rules

[22]. The complete LPP thesaurus is available online and a filter module can extract LPP references by querying keywords in the text label of the device. A *"STUDIO LPP list"* based on regular expressions was extracted from the French Health Insurance LPP thesaurus (consulted online in January 2022) to identify any THKSA-related IMD (Table 1). All the LPP references with a text label containing one of the key words "HIP", "KNEE" or "SHOULDER", but without "CUSHION", "PROTECTOR" or "CORRECTION", were included in the *STUDIO LPP list*. The number (and proportion) of LPP references finally included in the *STUDIO LLP list* was estimated by arthroplasty.

From the 21,699 LPP codes registered by the health insurance thesaurus, a total of 1,523 distinct LPP codes were included in the *STUDIO LPP list* for inclusion in the THKSA cohort (Table 1). The majority of LPP codes pertained to IMDs used in THA (51.1%).

Hospital discharge database

The hospital discharge database records the French CPT codes, known as CCAM. These CPT codes for THKSA were categorized by joint type and surgical indication, as follows: native joint replacement (*"remplacement de l'articulation native" in French*), one-stage IMD revision (*"changement de prothèse"*), 2-stage IMD revision (*"repose de prothèse"*) or revision (*"reintervention pour changement de tout ou partie de la prothèse"*). Only primary arthroplasties were included in the e-cohort (i.e. native joint replacement).

Mismatch between the 2 data sources: pharmacy/hospital discharge

A surgical hospital stay was identified as a "CPT but no LPP" mismatch when no LPP code from the *STUDIO LPP list* was recorded (i.e.: no LPP code at all or any other LPP reference). A stay was identified as a "LPP but no CPT" mismatch when no replacement CPT code was recorded (i.e.: no CPT code at all or any other CPT code). None of the mismatch hospital stays were included in the e-cohort. The number of patients affected by mismatches were detailed by type of arthroplasty and by hospital. A review of the stays with mismatches was performed to identify the main discrepancies between the two knowledge databases.

Analyses of the e-cohort, patients and stays

The absolute frequency of hospital stays (and patients) recorded in the pharmacy database involving the dispensation of IMD from the *STUDIO LPP list* of references was detailed by centre and joint type.

Similarly, the absolute frequency of stays (and patients) recorded in the hospital discharge database involving a primary THKSA was described by centre and joint type. Additionally, the proportion of hospital stays with an LPP code from the *STUDIO LPP list* but with a CPT code other than "replacement" was estimated, based on the discordant CPT codes recorded during those stays.

The number of hospital stays and patients included in the e-cohort was estimated by joint location (THKSA) and by care centre (I and II). The sex ratio and mean age with standard deviation (SD) were estimated, globally and by centre and joint. The main comorbidities were analysed using ICD-10 codes of all hospital stays in the hospital discharge database up to 2 years prior to the surgery of interest. The absolute (and relative) frequency of hospital stays involving at least one comorbidity were estimated by care centre and joint. The relative frequency of osteoarthritis coded as a primary diagnosis was estimated by joint.

Capability of pharmacy knowledge database

To explore its effectiveness in describing device characteristics, textual LPP labels were analysed by natural language processing (NLP). Given their frequency, hip IMD characteristics were examined as a use case. The IMD category was extracted from the hip LPP labels ("NEW SHORT LABEL") using regular expressions. The number (and proportion) of hip IMD references was estimated by category. Information regarding material type and cement usage was also extracted from the LPP labels. The number of LPP references presenting either of these two pieces of data was estimated by category.

Data management and processing was performed with R software (version 4 - R Core Team (2019)) [23]. R Foundation for Statistical Computing, Vienna, Austria. URL Project home page: https://cran.r-project.org/. Operating system(s): R studio server deployed on ubuntu server; Other requirements: R 3.6.0.

Table 1 STUDIO LPP list references of distinct services and products thesaurus (LPP) among the three joints

	Hip	Knee	Shoulder	Overall
LPP references included, n (%)	779 (51.1)	586 (38.5)	158 (10.4)	1,523 (100)

LPP: List of products and services

Ethics approval and consent to participate

Research in compliance with the Helsinki Declaration.

According to French regulations, all patients were informed (leaflet, poster, website) of the potential reuse of their data for research purposes and could refuse to participate. No signed consent was required. Data from eHOP® CDW are de-identified and linked to a unique anonymous identifier. Non-interventional studies from the eHOP® CDW were approved by the French Data Protection Agency (CNIL; N° 2212853 and N° 2212496).

Access to linked de-identified data in the datawarehouse was performed in accordance with the French Reference Methodology procedure for retrospective studies on data reuse, declaration signed by the teaching hospital of Tours (number 4116221019, 2019), as regulated by the French Data Protection Board (Commission Nationale de l'Informatique et des Libertés, CNIL). According to French data regulations, consent or information of each patient included was not required to use the French de-identified data. No nominative, sensitive or personal data of patients have been collected.

Results

Construction of the e-cohort

In the pharmacy database, one IMD dispensation was identified for 13,713 stays with a reference from the *STUDIO LPP list* (Table 2). The stays involved mostly hip arthroplasties (60.9%). Centre II presented more THA stays identified in the pharmacy database (+859) than centre I. On the contrary, more TKA and TSA stays were identified in centre I (+74 and +250, respectively).

From the 22 CPT codes from the French hospital discharge database (Table A2), one replacement was identified for 11,411 hospital stays according to the hospital discharge databases (Table 3). The most frequent procedure was hip arthroplasty (60.6% of overall stays). Centre II was concerned by more THA stays than centre I (n_{II} = 3,925 Vs. n_I = 2,991), though conversely, more TKAs and TSAs were identified in centre I.

LPP-CPT mismatch

Among all the patients initially identified by the hospital discharge database with a procedure code (n=10,257), the vast majority (96.2%) also had a corresponding IMD implantation (i.e. on same joint) from the LPP thesaurus (Fig. 1). The review of the medical records involving the remaining "CPT but no LPP" stays showed that most did

Table 2 Hospital stays (and patients) between 2010 and 2019 associated with a pharmacy record of services and products thesaurus (LPP) from the STUDIO LPP list, by hip, knee or shoulder arthroplasty (THKSA) and centre

	THA	TKA	TSA	Total
Stays with STUDIO LPP list records, n (%)				
Overall	8,348 (60.9)	3,941 (28.7)	1,424 (10.4)	13,713 (100)
Center I	3,693 (56.5)	2,007 (30.7)	837 (12.8)	6,537 (100)
Center II	4,655 (64.9)	1,934 (26.9)	587 (8.2)	7,176 (100)
Patients STUDIO LPP list records, n (%)				
Overall	7,282 (61.6)	3,282 (27.8)	1,255 (10.6)	11,819 (100)
Center I	3,170 (56.9)	1,684 (30.2)	719 (12.9)	5,573 (100)
Center II	4,112 (65.8)	1,598 (25.6)	536 (8.6)	6,246 (100)

THA: Total Hip Arthroplasty; TKA: Total Knee Arthroplasty;
TSA: Total shoulder Arthroplasty; LPP: List of products and services

Table 3 Overall number (proportion) of stays between 2010 and 2019 and corresponding patients, associated with a hospital discharge code for replacement by hip, knee or arthroplasty (THKSA) and centre

		THA	TKA	TSA	Total
Stays with replacement CPT records, n (%)					
	Overall	6,916 (60.6)	3,188 (27.9)	1,307 (11.4)	11,411 (100)
	Center I	2,991 (55.9)	1,609 (30.1)	745 (13.9)	5,345 (100)
	Center II	3,925 (64.7)	1,579 (26.0)	562 (9.2)	6,066 (100)
Patients with replacement CPT records, n (%)					
	Overall	6,304 (61.5)	2,764 (26.9)	1,189 (11.6)	10,257 (100)
	Center I	2,737 (57.2)	1,390 (29.0)	658 (13.7)	4,785 (100)
	Center II	3,567 (65.2)	1,374 (25.1)	531 (9.7)	5,472 (100)

THA: Total Hip Arthroplasty; TKA: Total Knee Arthroplasty; TSA: Total shoulder Arthroplasty; CPT: Current Procedure Terminology
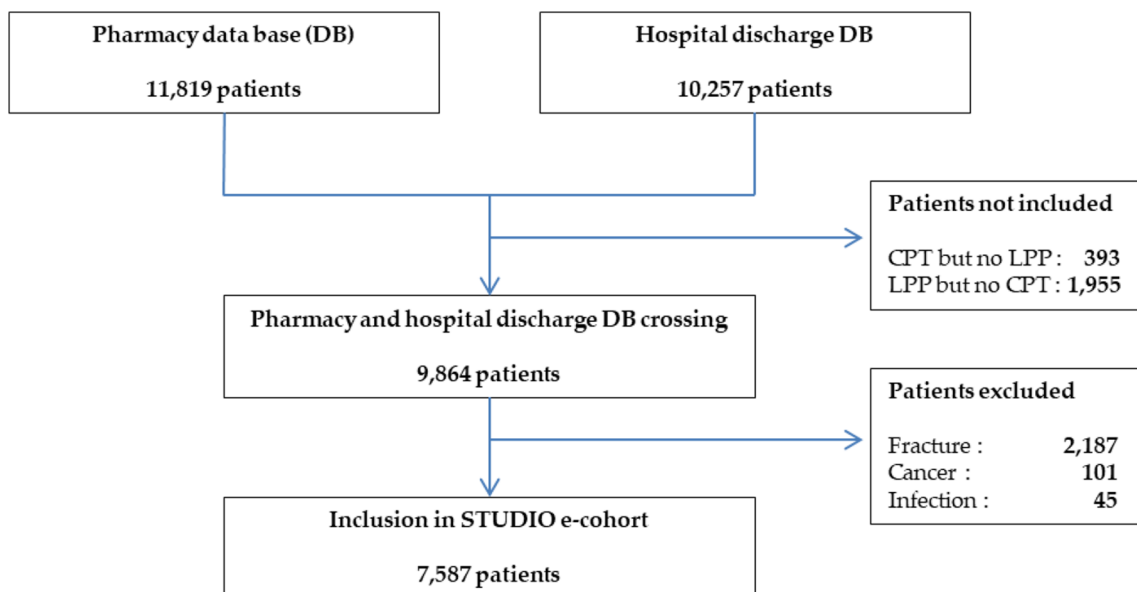

Fig. 1 Flow chart of patients included in the cohort, in both university hospitals for a first THKSA between 2010 and 2019 *DB: database; LPP: List of products and services; CPT: Current Procedure Terminology*

not involve THKSA, but another location of surgery (e.g. bone biopsy, fracture with immobilisation) or no arthroplasty at all (*Appendix table* A3).

Among the patients initially identified by the pharmacy database with IMD dispensation registered, most (83.4%) had also a corresponding procedure code (Fig. 1). The review of the medical records of the stays involving "LPP but no CPT" codes showed mismatches about the joint operated (e.g.: an LPP code for hip but a CPT code for knee) (*Appendix*, *Table* A3, *Figures* A2-4).

Patients were more frequently identified by the pharmacy source for THKSA (respectively + 978, + 518 and + 66) than by the hospital discharge database (Tables 3 and 4 and Figure A1). There was also a mismatch in the number of hospital stays that was constantly higher than the number of distinct patients, for each data source, centre and joint (Tables 3 and 4 and *Appendix Figure* A1).

e-cohort description

Between 2010 and 2019, 7,587 patients were included in the e-cohort for a first THKSA (Fig. 1). They represented 76.9% of the patients hospitalised over the period, with both records of a replacement arthroplasty (CPT code) and IMD dispensation of the *STUDIO LPP list* involving THKSA.

Among the patients identified by an IMD dispensation for THKSA, 16.5% (n=1,955) were not associated with any replacement procedure code during their hospital stay. Only 3.8% (n=393) of the patients initially identified by a procedure of interest in the hospital discharge database were not associated with any IMD reference from the pharmacy. The most frequent exclusion criterion was the presence of a fracture according to the ICD-10 coding during the arthroplasty hospital stay (n=2,187) involving 22.2% of the patients with THKSA (as identified by concordant data sources) in the center I whereas cancer appeared to be a more frequent diagnosis than fracture

for TKA for the centre II (Appendix, Figure A1). The e-cohort more frequently included patients from centre II for THA, but conversely, it involved fewer patients for TKA and TSA compared to center I (*Appendix*, *Table* A1).

The e-cohort involved 4,113 THA patients, 2,630 TKA patients and 844 TSA patients (Table 4). The overall sex ratio was 0.75 and the mean age 67.4 years, with a standard deviation of +/- 12.1. Obesity, cardio-vascular diseases and hypertension were the most frequent medical conditions whatever the location of the arthroplasty.

Obesity was more than three time lower in centre I (e.g.: 5.0% vs. 14.5% for the THA), as was hypertension (e.g.: 6.2% vs. 15.4% for TKA), whereas cancer was more prevalent (4.1% versus 1.9 for THA), as was diabetes mellitus. The most frequent diagnosis was arthrosis for the TKA (93.8%), THA (88.4%) and TSA (77.3%) cohorts.

Use case: hip IMD description

The 779 LPP references related to THA were categorized into 8 distinct categories corresponding to specific device parts using regular expressions. A ninth category, "other", was created for the least frequent references. Most of the THA LPP references involved the Stem (26%), Dual mobility (20%) and Head (14%) categories (Fig. 2).

Four IMD categories (Dual mobility, Head, Insert and Acetabular cup) had LPP references that specified the material type. Three IMD categories (Stem, Dual mobility and Acetabular cup) had LPP labels containing information about cement use with the device. None of the LPP references from the Cupula, Ring or Custom implant categories provided details on material or cement use. All LPP references in the Acetabular cup category presented information about material type, and the majority also mentioned cement use. Additionally, metrics information, such as diameter and size of the different IMD parts, was inconsistent and varied by IMD category or brand.

Table 4 Description of the cohort of patients hospitalised for a THA, TKA or TSA between 2010 and 2019

	THA		TKA		TSA	
	Centre I **(n=1,811)**	**Centre II** **(n=2,302)**	**Centre I** **(n=1,362)**	**Centre II** **(n=1,268)**	**Centre I** **(n=499)**	**Centre II** **(n=345)**
Sex ratio	0.99	0.81	0.80	0.69	0.49	0.41
Age, mean (SD)	66.8 (12.7)	65.5 (13.9)	69.1 (9.2)	68.6 (10.1)	70.1 (10.7)	68.6 (11.4)
Comorbidity, n(%)						
Cancer	75 *(4.1)*	27 *(1.2)*	27 *(1.9)*	6 *(0.5)*	8 *(1.6)*	1 *(0.3)*
Cardiovascular disease	127 *(7.0)*	159 *(6.9)*	85 *(6.2)*	82 *(6.5)*	24 *(4.8)*	14 *(4.0)*
Diabetes	41 *(2.3)*	88 *(3.8)*	35 *(2.6)*	68 *(5.4)*	5 *(1.0)*	16 *(4.6)*
Hypertension	111 *(6.1)*	275 *(11.9)*	85 *(6.2)*	196 *(15.4)*	31 *(6.2)*	51 *(14.8)*
Neurological disease	42 *(2.3)*	53 *(2.3)*	25 *(1.8)*	23 *(1.8)*	14 *(2.8)*	15 *(4.3)*
Overweight	91 *(5.0)*	335 *(14.5)*	104 *(7.6)*	320 *(25.2)*	26 *(5.2)*	62 *(17.9)*
Other comorbidity	732 *(40.4)*	1,235 *(53.6)*	587 *(43.1)*	799 *(63.0)*	164 *(32.9)*	173 *(50.1)*

THA Total Hip Arthroplasty; TKA: Total Knee Arthroplasty; TSA: Total shoulder Arthroplasty; CPT: Current Procedure Terminology
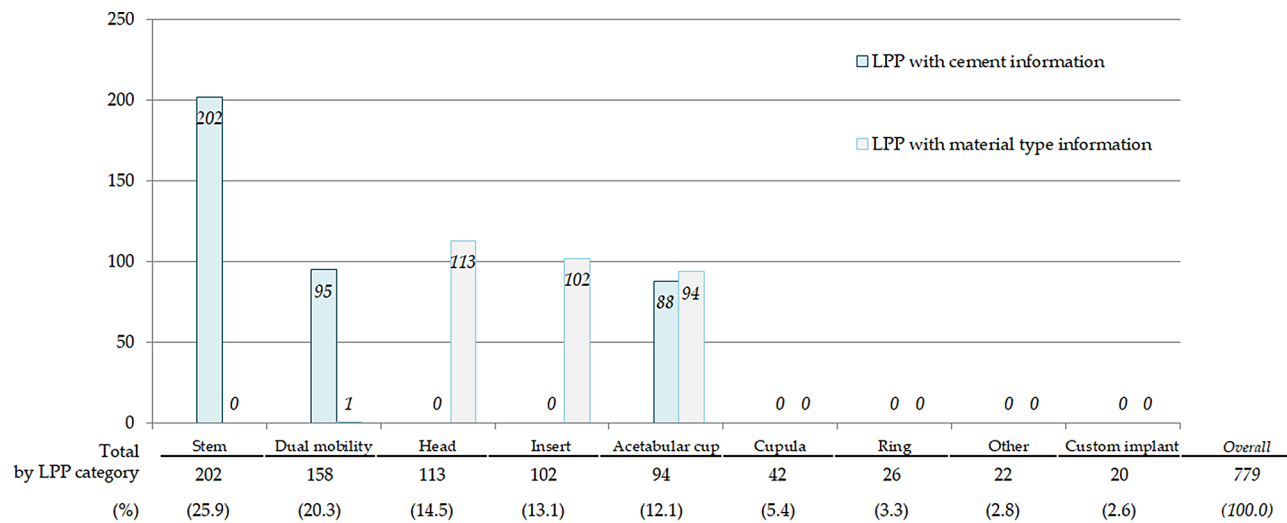
Fig. 2 LPP labels for Total Hip Arthroplasty Implantable Medical Devices, presenting textual information about IMD characteristics (cement or material), per device category Legend: Bars represent the number of LPP labels with cement or material type information in their text label from the LPP thesaurus, per IMD category; e.g.: all LPP labels from the stem category contain information about cement use with the device (n = 202) but any information about the material type

Discussion

This study demonstrated the advantages of using a CDW for device surveillance, using the monitoring of orthopaedic devices as a use case. It proved feasible even in the absence of a universal IMD knowledge database. This research successfully identified patients undergoing orthopaedic medical device implantation for native joint replacement crossing two data sources: hospital discharge (LPP domain terminology) and pharmacy databases (ICD-10). The e-cohorting method was initially developed and applied by the data center team at one center, and subsequently adapted and implemented at another center using the same CDW technology [19, 24]. Hence, the eHOP® data warehouse model used in each university hospital in Western France facilitated the straightforward adaptation and implementation of the e-cohorting script at a second hospital [19]. This approach improved the results and their reliability by increasing the number of patients and devices studied, along with completeness and quality of the registered information. Ultimately, implementing of an e-cohort for material surveillance will be easily workable across all eHOP® CDWs, which are expanding throughout France. Initially, we assumed that an comprehensive knowledge database could be used to describe IMDs (which are simply referenced by a code identifier or free text in the hospital information system). It turned out that no such database was currently fully satisfactory for the automatic annotation of MD references, but the LPP thesaurus was a good option to start. Our approach aims to close the gap between conventional epidemiological cohorting tools and the implementation of an automated e-cohort.

The device characteristics were not yet structured in eHOP®, but could have been extracted from the LPP label using natural processing techniques based on regular expressions. However, information stored in text is not homogeneous and its presence depends of the type of device. Further work on data structuring and convergence in the medical files is necessary, such as in surgical reports. A research work on that specific challenge has been undergone successfully using CDWs in the HAC-ROHUGORTHO study [25]. Including the LPP thesaurus and CLADIMED, the medical device interoperability catalogue (CIODM) aims to structure the information of more than 1.6 million medical devices [26]. EUDAMED terminology (replacing CLADIMED) is an IT system established on medical devices and developed by the European Commission [27] and could represent the next step to interoperability. Indeed, the registration of legacy devices in EUDAMED is not yet required as the system is not fully functional, but its application started in 2021. This implementation will ease the e-cohorting process, but our pilot study showed we could already move on in France, with the Health Insurance data as a huge legacy of medical big data. As recently shown in a literature review, data collection automation through the reuse of real-world digital data from clinical data warehouses represents a great opportunity to improve medical device monitoring [14, 28]. However, a number of obstacles remain, such as data quality and interoperability through the use of common and regularly updated terminologies, and the use of a Unique Device Identifier (UDI).

This approach had some limits. Considering both data sources (pharmacy and hospital discharges), more

hospital stays were identified than individual patients, likely due to contralateral arthroplasty being performed on the other joint side of the same patient. The patient profiles were consistent to those described in the literature for hip procedures in terms of age, gender, and comorbidities [29, 30]. Patients were not included in the study if a fracture diagnosis was coded during the hospital stay because the complication of those "after fracture arthroplasties" are very different. Comorbidity rates were estimated from the hospital discharge data with the international classification of disease codes (ICD-10). Their proportion depends on the centre's coding strategy. Regarding the hospital information department, the quality review could result in possible underestimation of comorbidities compared to registry studies [31]. For this reason, the estimation of the proportion of comorbidities according to the hospital discharge database might vary according to hospitals hosting the data warehouse and to the department of medical information's coding strategy. According to the hospital discharge database, CPT codes corresponding to THA were more frequent than TKA, as shown in the literature, but with a higher ratio [32]. Epidemiological grounds or centre strategy for the development of specific activity can explain the different number of patients between centres or arthroplasty. Indeed, university hospitals with data centres are not representative of all hospitals in France, as they tend to be more specialized and attract a larger patient population beyond their local communities. This likely explains the higher number of hip arthroplasty cases observed, possibly due to the higher incidence of hip dysplasia in this region of France [33]. Also, one centre developed its activity particularly towards shoulder arthroplasties, as a reference centre, explaining its higher proportion compared to the other university hospital.

Finally, a program was developed to automate information extraction for patients and devices using natural language processing (NLP) on pharmacy labels, due to the lack of an international IMD mapping system to describe prosthetic device characteristics and the absence of global and dynamic repositories in France. The resulting e-cohort included 7,587 patients who underwent initial THKSA. Compared to traditional device surveillance models, utilizing real-world digital health data offers a more effective way to monitor joint replacement surgeries by identifying target populations, population size, and key descriptive results. Further studies are needed to implement machine learning models for NLP of medical records, such as surgery reports, to extract critical information like the laterality of the procedure, as a recent study proposed [25]. Automatic and dynamic information availability from electronic health data concerning IMD opens opportunities for dynamic real-world monitoring to assess associated risks related to implanted materials [25, 28] and enhance feedbacks to orthopaedists surgeons and infectious disease specialists about implant follow-up, guiding them in therapeutic decision-making, and inform public health policymakers [4, 8, 14, 25, 34]. This study demonstrated the interest of artificial intelligence tools in healthcare for efficient and cost-effective automated surveillance, in compliance with recent European regulations [17]. Matching the hospital database to the national health data system (SNDS) could help developing a shareable model for extra-hospital material health surveillance.

Conclusion

This pilot study demonstrated the feasibility of an e-cohort of orthopaedic IMD involving native joint replacement. An automation script for information extraction from structured data was implemented, adapted and tested in two centres to describe THKSA patients and device characteristics, showing the easy transferability of the script to hospitals with a clinical data warehouse. The model could be developed in other CDW centres. Intra-hospital follow-up can be conducted to describe the infectious, thromboembolic and unsealing complications. Automation of information extraction concerning the IMD and patient's features and follow-up will participate in the construction of the new material-vigilance era.

Summary points

- This pilot study demonstrated the feasibility of an e-cohort of orthopaedic devices using clinical data warehouses.
- Intra-hospital follow-up can be conducted to describe the infectious, thromboembolic and unsealing complications.

Supplementary Information

Supplementary Material 1

Supplementary Material 2

Author contributions
Conceptualisation, P.Rosset, L.Grammatico-Guillon; Methodology, P.Rosset, C.Salpétrier, J.Herbert, M.Cuggia, L.Grammatico-Guillon; Validation and formal analysis, C. Salpétrier, J. Herbert and M.Ansoborlo. Writing—original draft preparation, M.Ansoborlo; Writing—review and editing, C.Salpétirer, J. Herbert P.Rosset, L.Grammatico-Guillon, Louis-Romée Le Nail and M.Cuggia; Funding acquisition, L.Grammatico-Guillon and P. Rosset. All authors have read and agreed to the published version of the manuscript.

Data availability
The data presented in this study are available on request from the corresponding author.

Declarations

Ethics approval and consent to participate
Not applicable According to French regulations, all patients were informed (leaflet, poster, website) of the potential reuse of their data for research purposes and could refuse to participate. No signed consent was required. Data from eHOP® CDW are de-identified and linked to a unique anonymous identifier. Non-interventional studies from the eHOP® CDW were approved by the French Data Protection Agency (CNIL; N° 2212853 and N° 2212496). Access to linked de-identified data in the datawarehouse was performed in accordance with the French Reference Methodology procedure for retrospective studies on data reuse, declaration signed by the teaching hospital of Tours (number 4116221019, 2019), as regulated by the French Data Protection Board *(Commission Nationale de l'Informatique et des Libertés, CNIL)*. According to French data regulations, consent or information of each patient included was not required to use the French de-identified data. No nominative, sensitive or personal data of patients have been collected.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

References
1. Putman S, Girier N, Girard J, Pasquier G, Migaud H, Chazard E. Épidémiologie des prothèses de hanche en France: analyse de la base Nationale Du PMSI De 2008 à 2014. Revue De Chirurgie Orthopédique et Traumatologique. 2017;103:S90.
2. Erivan R, Tardieu A, Villatte G, Ollivier M, Jacquet C, Descamps S et al. Knee surgery trends and projections in France from 2008 to 2070. Orthopaedics & Traumatology: Surgery & Research. 2020;106:893–902.
3. Villatte G, Erivan R, Barth J, Bonneviale N, Descamps S, Boisgard S. Progression and projection for shoulder surgery in France, 2012–2070: Epidemiologic study with trend and projection analysis. Orthopaedics & Traumatology: Surgery & Research. 2020;106:1067–77.
4. Grammatico-Guillon L, Miliani K, Banaei-Bouchareb L, Solomiac A, Sambour J, May-Michelangeli L et al. A computerized indicator for surgical site infection (SSI) assessment after total hip or total knee replacement: the french ISO-ORTHO indicator. Infect Control Hosp Epidemiol. 2022;43(9):1171–8. https://doi.org/10.1017/ice.2021.371. Epub 2021 Sep 9. PMID: 34496983.
5. IQSS. 2021 - Évènements thrombo-emboliques après pose de prothèse totale de hanche (ETE-PTH). Haute Autorité de Santé. https://www.has-sante.fr/jcms/p_3293932/en/iqss-2021-evenements-thrombo-emboliques-apres-pose-de-prothese-totale-de-hanche-ete-pth. Accessed 31 May 2022.
6. IQSS. 2021 - Infections du site opératoire après pose de prothèse totale de hanche (ISO-PTH). Haute Autorité de Santé. https://www.has-sante.fr/jcms/p_3294825/en/iqss-2021-infections-du-site-operatoire-apres-pose-de-prothese-totale-de-hanche-iso-pth. Accessed 31 May 2022.
7. Le Manach Y, Collins G, Bhandari M, Bessissow A, Boddaert J, Khiami F, et al. Outcomes after hip fracture surgery compared with elective total hip replacement. JAMA. 2015;314:1159–66.
8. Grammatico-Guillon L, Rusch E, Astagneau P. Surveillance of prosthetic joint infections: international overview and new insights for hospital databases. J Hosp Infect. 2015;89:90–8.
9. Yao JJ, Hevesi M, Visscher SL, Ransom JE, Lewallen DG, Berry DJ, et al. Direct Inpatient Medical costs of Operative Treatment of Periprosthetic hip and knee infections are twofold higher than those of aseptic revisions. J Bone Joint Surg. 2021;103:312–8.
10. Knebel C, Menzemer J, Pohlig F, Herschbach P, Burgkart R, Obermeier A, et al. Periprosthetic joint infection of the knee causes high levels of Psychosocial Distress: a prospective cohort study. Surg Infect. 2020;21:877–83.
11. Kunutsor SK, Whitehouse MR, Blom AW, Beswick AD, INFORM Team. Patient-related risk factors for Periprosthetic Joint Infection after total joint arthroplasty: a systematic review and Meta-analysis. PLoS ONE. 2016;11:e0150866.
12. Lindgren JV, Wretenberg P, Kärrholm J, Garellick G, Rolfson O. Patient-reported outcome is influenced by surgical approach in total hip replacement: a study of the Swedish hip Arthroplasty Register including 42,233 patients. Bone Joint J. 2014;96–B:590–6.
13. Courtney PM, Markel DC. Arthroplasty registries: improving clinical and economic outcomes. J Knee Surg. 2017;30:7–11.
14. Bautista MP, Bonilla GA, Mieth KW, Llinás AM, Rodríguez F, Cárdenas LL, et al. Data Quality in Institutional Arthroplasty registries: description of a model of validation and report of preliminary results. J Arthroplasty. 2017;32:2065–9.
15. Niederländer C, Wahlster P, Kriza C, Kolominsky-Rabas P. Registries of implantable medical devices in Europe. Health Policy. 2013;113:20–37.
16. Delaunay C. Registries in orthopaedics. Orthop Traumatology: Surg Res. 2015;101:S69–75.
17. Niemiec E. Will the EU Medical device regulation help to improve the safety and performance of medical AI devices? Digit Health. 2022;8:20552076221089080.
18. van Mourik MSM, van Rooden SM, Abbas M, Aspevall O, Astagneau P, Bonten MJM, et al. PRAISE: providing a roadmap for automated infection surveillance in Europe. Clin Microbiol Infect. 2021;27(Suppl 1):S3–19.
19. Madec J, Bouzillé G, Riou C, Van Hille P, Merour C, Artigny M-L, et al. eHOP Clinical Data Warehouse: from a prototype to the creation of an Inter-regional Clinical Data Centers Network. Stud Health Technol Inf. 2019;264:1536–7.
20. Cuggia M, Guillon-Grammatico L, Gourraud P-A, Chretien J-M, Bocquet F, Guiton V, et al. The Ouest Data Hub: an Interregional Health Data Sharing Ecosystem for Research. Stud Health Technol Inf. 2024;316:1679–83.
21. Dhalluin T, Fakhiri S, Bouzillé G, Herbert J, Rosset P, Cuggia M, et al. Role of real-world digital data for orthopedic implant automated surveillance: a systematic review. Expert Rev Med Dev. 2021;18:799–810.
22. Décret. n° 2019–571 du 11 juin 2019 relatif à l'identification individuelle des produits et prestations inscrits par description générique sur la liste prévue à l'article L. 165-1 du code de la sécurité sociale - Légifrance. https://www.legifrance.gouv.fr/loda/id/JORFTEXT000038578527/. Accessed 8 Sep 2022.
23. R: The R Project for Statistical Computing. https://www.r-project.org/. Accessed 3 Jun 2022.
24. Bouzillé G, Westerlynck R, Defossez G, Bouslimi D, Bayat S, Riou C, et al. Sharing Health Big Data for Research - A Design by Use cases: the INSHARE platform Approach. Stud Health Technol Inf. 2017;245:303–7.
25. Ansoborlo M, Cardoso J, Herbert J, Salpetrier C, Bouzille G, Cuggia M, et al. Performance of a NLP Tool for text classification from Orthopaedic Operative reports, using data from the Large Network of Clinical Data Warehouses of the West of France: the HACRO-HUGORTHO project. Stud Health Technol Inf. 2024;316:1979–83.
26. CIOdm. May, le référentiel d'interopérabilité du dispositif médical – PHAST. https://www.phast.fr/ciodm/. Accessed 30 2022.
27. UDI/Devices registration. https://ec.europa.eu/health/medical-devices-eudamed/udidevices-registration_en. Accessed 31 May 2022.
28. Dhalluin T, Fakhiri S, Bouzillé G, Herbert J, Rosset P, Cuggia M, et al. Role of real-world digital data for orthopedic implant automated surveillance: a systematic review. Expert Rev Med Devices. 2021;18:799–810.
29. Giori NJ, Radin J, Callahan A, Fries JA, Halilaj E, Ré C, et al. Assessment of Extractability and Accuracy of Electronic Health Record Data for Joint Implant Registries. JAMA Netw Open. 2021;4:e211728.
30. Pivec R, Johnson AJ, Mears SC, Mont MA. Hip arthroplasty. Lancet. 2012;380:1768–77.
31. Mussa M, Manciulli T, Corbella M, Mariani B, Cambieri P, Gipsz N, et al. Epidemiology and microbiology of prosthetic joint infections: a nine-year, single-center experience in Pavia, Northern Italy. Musculoskelet Surg. 2021;105:195–200.
32. Bayliss L, Culliford D, Monk A, Glyn-Jones S, prieto-alhambra D, Judge A et al. The effect of patient age at intervention on risk of implant revision after total replacement of the hip or knee: a population-based cohort study. Lancet. 2017;389.
33. Masse A. [History and epidemiology of congenital hip dislocation in Brittany]. Acta Orthop Belg. 1990;56(1 Pt A):43–52.
34. Grammatico-Guillon L, Baron S, Gaborit C, Rusch E, Astagneau P. Quality assessment of hospital discharge database for routine surveillance of hip and knee arthroplasty-related infections. Infect Control Hosp Epidemiol. 2014;35:646–51.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.