Challenges of Trustable AI and Added-Value on Health
B. Séroussi et al. (Eds.)

doi:10.3233/SHTI220447

Automated Coding in Case Mix Databases of Bacterial Infections Based on Antimicrobial Susceptibility Test Results

Radia SPIGA[a,b], François-Elie CALVIER[a], Anne CARRICAJO[c], Bruno POZZETTO[d], Béatrice TROMBERT-PAVIOT[a] and Cédric BOUSQUET[a,e, 1]
[a] *Public health and medical information unit, Saint Etienne University Hospital, France*
[b] *Mines Saint-Etienne, Univ Clermont Auvergne, CNRS, UMR 6158 LIMOS, Institut Henri Fayol, F – 42023*
[c] *Laboratoire des Agents Infectieux et d'Hygiène, Centre Hospitalo-Universitaire de Saint-Etienne, Saint-Etienne, France*
[d] *CIRI (Centre International de Recherche en Infectiologie), Université de Lyon, Université Claude Bernard Lyon 1, INSERM U1111, CNRS, UMR5308, ENS Lyon, Université Jean Monnet de Saint-Etienne, Lyon, France.*
[e] *Sorbonne Université, INSERM, Univ Paris 13, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances pour la e-Santé, LIMICS, 75006, Paris, France*

Abstract. Our objective was to improve the accuracy of bacteria and resistance coding in a hospital case mix database. Data sources consisted of 50,074 files on bacteriological susceptibility tests transmitted with the HPRIM protocol from laboratory management system to electronic health record of the University hospital of Saint Etienne in July 2017. An algorithm was implemented to detect susceptibility tests containing information corresponding to codes whose addition in the case mix database was susceptible to increase the severity level of a diagnosis related group. Among 132 hospital stays fulfilling the conditions, 27 were lacking bacteria and/or resistance codes, and the tariff was increased for 9 stays, with earnings of €54,612. Analyzing Antimicrobial susceptibility tests helps to improve clinical coding and optimize the financial gain.

Keywords. Clinical Coding, Electronic Health Records, Bacteria, Health care payment system, France

1. Introduction

Funding of acute care in French hospitals is based on Prospective Payment System (PPS) that relies on the description of care activity. It is evaluated through a Diagnosis Related Group (DRG)-based information system called PMSI (*Programme de Médicalisation du Système d'Information* for "medicalized information system program"). This program uses the 10th revision of the International statistical Classification of Diseases and related health problems (ICD-10). A Patient Classification System is applied on the PMSI data to assign a hospital stay to a DRG. Each DRG has a tariff set by the Ministry of Health

---

[1] Corresponding Author, Dr Cedric Bousquet, SSPIM, Bâtiment CIM42, chemin de la Marandière, Hôpital Nord, 42055 Saint Etienne, France; E-mail: cedric.bousquet@chu-st-etienne.fr.

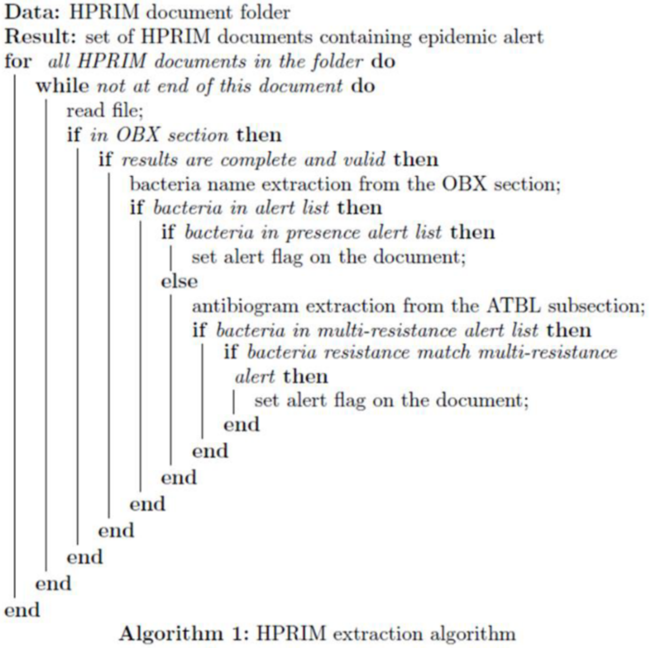
that varies according to the most responsible diagnosis and the severity scale from 1 to 4, corresponding to the severity of Comorbidities or Complications (CCs).

The PMSI database (French case-mix database) is commonly used for epidemiologic research and surveillance. However, several studies have pointed the low accuracy of associated diagnoses in the different hospital case-mix databases available worldwide, especially in case of bacterial infection [1–4]. In France, codes specifying the bacterium responsible for an infection or its resistance to certain drugs belong to the list of the CCs whereas codes related to highly-resistant bacteria induce an increased level of severity to 3 or 4.

We present here a pilot study using files sent from laboratory management system (LMS) to patients' electronic health record (HER). This file exchange follows the French HPRIM protocol (*Harmoniser et PRomouvoir l'Informatique* Médicale for "standardize and promote medical information technology") which allows a secure electronic transmission of biological results. We developed an algorithm for analyzing HPRIM files and targeting those responding to established criteria with the objective of demonstrating the feasibility to improve the coding of bacteria and resistance in a case mix database.

2. Methods

The results of susceptibility tests of bacteria from the Department of Infectious Agents and Hygiene of the University-Hospital of Saint-Etienne are stored in HPRIM files together with other data including the sampling site, the type of culture and the bacterium species. HPRIM files produced during the month of July 2017 by the LMS were extracted and stored safely on a server. At the time of extraction all the infectious agents were processed, including bacteria, viruses and parasites, without distinction.



```
Data: HPRIM document folder
Result: set of HPRIM documents containing epidemic alert
for all HPRIM documents in the folder do
    while not at end of this document do
        read file;
        if in OBX section then
            if results are complete and valid then
                bacteria name extraction from the OBX section;
                if bacteria in alert list then
                    if bacteria in presence alert list then
                        set alert flag on the document;
                    else
                        antibiogram extraction from the ATBL subsection;
                        if bacteria in multi-resistance alert list then
                            if bacteria resistance match multi-resistance
                              alert then
                                set alert flag on the document;
                            end
                        end
                    end
                end
            end
        end
    end
end
```

Algorithm 1: HPRIM extraction algorithm

Figure 1. Algorithm implemented to extract susceptibility tests from HPRIM files

We identified the parts of HPRIM format that describe antimicrobial susceptibility tests. We implemented a program using the Java language for reading the HPRIM files and selecting susceptibility tests related to bacteria and corresponding to a valid result, which excluded non-bacterial agents and invalid susceptibility tests figure 1). Bacteriological results corresponding to short hospital stays (less than 2 nights) were also excluded.

Then we targeted bacteria and resistance profiles corresponding to those whose presence leads to increase the severity level of the DRG (codes with severity levels of 3 and 4): the selected bacteria were *Staphylococcus aureus*, *Proteus mirabilis*, *Pseudomonas aeruginosa*, *Bacillus fragilis*, *Clostridium difficile* and *Clostridium perfringens*; the selected resistance profiles were methicillin-resistant *Staphylococcus aureus* (MRSA), extended-spectrum beta-lactamase-producing *Enterobacteriaceae* (ESBL-PE), vancomycin-resistant Enterococci (VRE) and carbapenemase-producing *Enterobacteriaceae*.

To code validly bacteria and antimicrobial resistance, the following PMSI regulatory conditions must be fulfilled: presence of a susceptibility test, mention of the infection on the EHR, and prescription of an antimicrobial treatment. For every hospital stay containing a susceptibility test according to the selection criteria, the accuracy of the ICD-10 coding in the hospital case-mix database was verified, with two possible scenarios: (i) in the absence of ICD-10 codes of bacteria or resistance, the EHR was checked to find information allowing the coding update; (ii) in its presence, the EHR was verified to find arguments justifying the coding. Then the sensitivity and specificity of initial bacterium coding were calculated. Finally, when missing ICD-10 codes of bacteria and/or resistance were added, the financial gain corresponding to the coding modification was calculated.

3. Results

From the 50,074 laboratory results performed by the laboratory on infectious agents during the month of July 2017, 159 susceptibility tests fulfilling the selection criteria were realized, corresponding to 132 hospital stays (Figure 2). After checking the case-mix database and the EHR, bacteria and antimicrobial resistance were appropriately coded for 20 hospital stays.

For the remaining stays without bacteria and/or resistance codes, conditions for adding such code were present for 27 stays (Table 1). Initial coding of bacteria and/or resistance in the PMSI database had a sensitivity of 42.5% (95% CI 28.26 - 57.82) and a specificity of 97.6% (95% CI 91.76 - 99.71). Among the 27 hospital stays that fulfilled an upgrading of coding conditions, the tariff of 9 stays was significantly increased, with overall earnings of €54,612. Among stays with unfulfilled coding conditions, two stays were wrongly coded; codes were removed without financial consequence.

Table 1. Comparative results of two PMSI coding strategies relative to 132 hospital stays including a significant bacterial infection with susceptibility testing, recorded in July 2017 at the University-Hospital of Saint-Etienne, France.

	Retrospective coding: Infection	Retrospective coding: No Infection	Total
Initial coding: Infection	20	2	22
Initial coding: No infection	27	83	110
Total	47	85	132

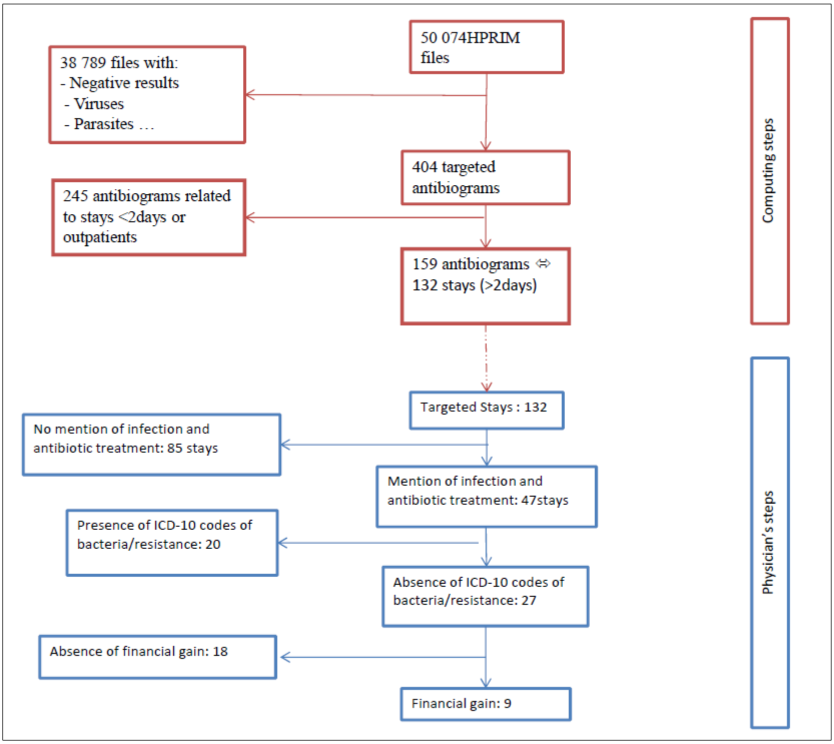


Figure 2. Study steps flow chart (Red: Antimicrobial susceptibility tests selecting algorithm; Blue: Description of checking and improving coding steps)

4. Discussion

With a sensitivity of 42.5%, coding bacteria and resistance remains incomplete, which is consistent with a study carried on the 2014 French case-mix database that showed nationwide inconsistencies between the type of infections coded and the associated bacteria and resistance. Although only 9 out of 132 stays have been upgraded, the recovered benefit was important, corresponding for instance to the total employer cost of a nurse or a technician over one year.

Our results show a high specificity of initial coding in the PMSI database since only 2 stays were incorrectly coded, whose deletion of the corresponding codes led to no financial loss. In addition to searching for missing codes, we also aim to check codes already present. Indeed, removing incorrect codes allows avoiding financial penalties from the national Health Insurance resulting of controls of the consistency between diagnoses in the case-mix database, and information available on the medical records.

Usually, predicting PMSI coding exploits supervised machine learning methods over unstructured data like medical observation. Nevertheless, Scheurwegs et al. used structured data such as biological test results and drug prescription to improve the machine learning results [5]. However, machine learning methods identify data co-occurrences, which could lead to association rules that do not necessarily reflect expert knowledge. The decision rules implemented in our approach were defined from the regulatory ones established in France by the national Health Insurance, which makes them intuitive to understand.

To our knowledge, this is the first time that a proof of concept is provided for supporting the efficiency and the return on investment of implementing automated analysis of available electronic data to improve coding of bacterial infection in a case-mix database. Several works that involved different data sources such as microbiology, pharmacy and case mix database have been published [6,7] about the surveillance of healthcare associated infections. However, emphasis was on surveillance, and not on improvement of coding. Clinical decision support systems have previously supported antimicrobial stewardship programs by collating data from multiple sources, including microbiology and pharmacy [8]. Part of the processing of these systems is based on the analysis of new culture results when they become available, which results in the generation of alerts regarding resistant pathogens. However, it differs from our approach as it intends to improve both antimicrobial prescription and appropriateness. As for electronic surveillance systems, it was acknowledged that efficient processing was dependent on their ability to interface with institutional data sources.

We propose to work directly on messages transmitted by the LMS containing the antimicrobial susceptibility tests rather than extracting this information from a database. This strategy requires a complete interoperability between the LMS and the coding tool. In our hospital, such messages are based on the HPRIM format. This format is not a standard at the international level but is close to HL7 V2.x that is implemented in many hospital information systems worldwide. For this reason, we are confident that a similar approach could be implemented in other clinical health information systems, even if they do not use the HPRIM protocol.

References

[1] Gedeborg R, Furebring M, Michaëlsson K. Diagnosis-dependent misclassification of infections using administrative data variably affected incidence and mortality estimates in ICU patients. J Clin Epidemiol. 2007;60(2):155.e1-e155.e11. doi:10.1016/j.jclinepi.2006.05.013

[2] Pinyo Rattanaumpawan MD, Wongkamhla T, Thamlikitkul V. Accuracy of ICD-10 coding system for identifying comorbidities and infectious conditions using data from a Thai University Hospital administrative database. J Med Assoc Thai. 2016;99(4):368-373.

[3] Wilhelms SB, Walther SM, Huss F, Sjöberg F. Severe sepsis in the ICU is often missing in hospital discharge codes. Acta Anaesthesiol Scand. 2017;61(2):186-193. doi:10.1111/aas.12814

[4] Chan M, Lim PL, Chow A, Win MK, Barkham TM. Surveillance for Clostridium difficile Infection: ICD-9 Coding Has Poor Sensitivity Compared to Laboratory Diagnosis in Hospital Patients, Singapore. Braga EM, ed. PLoS ONE. 2011;6(1):e15603. doi:10.1371/journal.pone.0015603

[5] Scheurwegs E, Luyckx K, Luyten L, Daelemans W, Van den Bulcke T. Data integration of structured and unstructured sources for assigning clinical codes to patient stays. J Am Med Inform Assoc. 2016;23(e1):e11-e19. doi:10.1093/jamia/ocv115

[6] de Bruin JS, Seeling W, Schuh C. Data use and effectiveness in electronic surveillance of healthcare associated infections in the 21st century: a systematic review. J Am Med Inform Assoc. 2014;21(5):942-951. doi:10.1136/amiajnl-2013-002089

[7] Freeman R, Moore LSP, García Álvarez L, Charlett A, Holmes A. Advances in electronic surveillance for healthcare-associated infections in the 21st Century: a systematic review. J Hosp Infect. 2013;84(2):106-119. doi:10.1016/j.jhin.2012.11.031

[8] Forrest GN, Van Schooneveld TC, Kullar R, Schulz LT, Duong P, Postelnick M. Use of Electronic Health Records and Clinical Decision Support Systems for Antimicrobial Stewardship. Clin Infect Dis. 2014;59(suppl_3):S122-S133. doi:10.1093/cid/ciu565