HAL open science

Using SNOMED-CT to Help the Transition from Microbiological Data to ICD-10 Sepsis Codes.

Iris Ternois, Typhaine Billard-Pomares, Etienne Carbonelle, Loriane Franchinard, Catherine Duclos

To cite this version:

Iris Ternois, Typhaine Billard-Pomares, Etienne Carbonelle, Loriane Franchinard, Catherine Duclos. Using SNOMED-CT to Help the Transition from Microbiological Data to ICD-10 Sepsis Codes.. Studies in Health Technology and Informatics, 2019, 264, pp.1604-1605. 10.3233/SHTI190556 . hal-03479359

HAL Id: hal-03479359

https://hal.science/hal-03479359v1

Submitted on 16 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)
© 2019 International Medical Informatics Association (IMIA) and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/SHTI190556

Using SNOMED-CT to Help the Transition from Microbiological Data to ICD-10 Sepsis Codes

Iris Ternois[a], Typhaine Billard-Pomares[b], Etienne Carbonelle[b], Loriane Franchinard[c], Catherine Duclos[a,b]

[a] Univ Paris 13, Sorbonne Université, INSERM, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances pour la e-Santé, LIMICS, F-93019 Bobigny, France
[b] Hôpital Avicenne, AP-HP, Bobigny, France
[c] Foetal Medicine Department, Hôpital Armand Trousseau, AP-HP, Paris, France

Abstract

Assigning ICD-10 code of sepsis in regard of a pathogenic bacterium found in an haemoculture requires knowledge of microbiology because of the difference of granularity. The aim of this paper is to automate this coding thanks to the use of SNOMED-CT. A dichotomous classification of bacteria causing sepsis has been generated in respect of ICD-10. Our algorithm follows this and explores SNOMED-CT to assign the right ICD-10 code of the sepsis. Applied to a list of 164 bacteria, the system has an error rate of 1.22 %.

Keywords:

Medical coding, International Classification of Diseases, Systematized Nomenclature of Medicine

Introduction

Using medical terminologies to code health care data is an essential process, it allows the standardisation of this data, which is a necessary condition of their re-use in computer processings. The coding procedure is complex and it is difficult to achieve proper coding in terms of completeness or accuracy [1].

The cognitive process to code often implies to conciliate natural language data in clinical documentation and terminology concepts. It requires medical knowledge to find synonyms, to infer aggregation or specialization reasonings, and to get acquainted with the terminology usage rules [2,3].

In the field of infectiology, the available data is often a precise and textual description of the infection, with the identification of the pathogenic agent, easily trackable in lab results. In order to match this description into ICD-10 codes, there is a necessary operation to switch from a thin granularity (bacteria name) to a larger granularity (bacteria group).

To infer this transformation, a knowledge mobilization is necessary. This knowledge may be unavailable, or forgotten. In the domain of microbiology, a reasonning mediator may help to achieve this task [4] . For exemple, to assign the ICD-10 code A415 (Gram negative sepsis) from the result of a positive blood-culture containing Escherichia coli, knowledge about gram coloration is needed.

Systematized Nomenclature of Medicine - Clinical Terms (SNOMED-CT) terminology covers every field of medicine and depicts bacteria into multi point-of-view hierarchies (cultural characteristic, taxonomy…). The hypothesis is that SNOMED-CT can create the necessary pivot to infer the correct affiliation to a bacteria to a more generic group as described in ICD-10.

The aim of this study is to developp a method to assign the right ICD-10 code for a sepsis, based on the name of the bacteria that caused it, found in the bacteriological results.

Methods

Constitution of a Gold Standard

A list of bacteria was created using an anonymous and agregated collection of positive blood-cultures extracted from the microbiological lab of Avicenne Hospital between January and July 2018. The positive blood-cultures not associated with any antibiotic resistance testing has been excluded. The list is composed of 164 distinct bacteria.

For each bacteria, two microbiologists gave an ICD-10 code for the potentially associated sepsis. The coding was later checked by a specialist in medical information coding. The discrepancies have been jointly corrected. This final list of ICD-10 codes constitutes the Gold Standard.

Construction of a classification of bacteria causing a sepsis according to ICD-10

The classification of spesis causing bacteria follows the logic behind the ICD-10 classification of sepsis. The two ICD-10 classes of interest are A40 (Streptococcal septicaemia) and A41 (Others septicaemia). They describe most cases of sepsis, except some exclusions.

The ICD-10 classification of these two categories have been studied, along with all the exclusions in order to build a dichotomous decision tree for the speis causing bacteria. Every leaf of the tree correspond to a ICD-10 code. This dichotomization allows to push as the last class the unprecise codes (A 408 : "Other streptococcal sepsis" or A418 : "Other specified sepsis"). Furthermore, the code A419 : "Sepsis, unspecified" is excluded because in our study the bacteria are identified from the blood sample, thus the "unspecified" code doesn't apply here.

ICD-10 sespsis coding algorithm

The principle is to find for each bacteria found in blood cultures sufficient information to map to the bacterial point of view of ICD-10 previously described.

For each bacteria tested, the algorithm (developed in Python) browses the preceding tree by verifying the matching with the leaf. There are three different situations.

Codes mentioning a bacteria defined by its species

The leaf describes a named bacteria. For instance, the code A410 is sepsis due to staphylococcus aureus. When the algorithm comes to the leaf "Staphylococcus aureus", it checks if the bacteria is a staphylococcus aureus, and assigns the code A410 if it is the case. This situation is the simplest, the assignation of the code is direct.

Codes mentioning a disease

The exclusions of the classes A40 and A41 can mention diseases associated with sepsis without the name of pathogenic bacteria (e.g. "Sepsis in tularemia"). SNOMED-CT is here used to find the causative agents of diseases. The PyMedTermino [5] module allows to explore these relationships. When the bacteria tested is part of the causative agents, the ICD-10 code is attributed.

Codes mentioning a bacteria defined by a grouping (genus, resiratory mode, etc.)

The leaf describes a group of bacteria. For instance, the code A415 is sepsis due to gram-negative bacteria. In order to check if the bacteria is a gram-negative bacillus, the algorithm exploration is managed with the PyMedTermino module.

First, the SNOMED-CT concept associated with the bacteria tested is retrieved. Then, the list of the parents of this concept is searched. The algorithm verifies the presence of the name of name of the group (e.g. "gram-negative") is found, the algorithm stops at this leaf. Else, the algorithm goes to the next leaf.

Evaluation

The results of the algorithm for the 164 bacteria have been confronted to the Gold Standard. Any difference of coding is a wrong assignation of the code by the algorithm.

Results

The 164 bacteria of the Gold Standard are split into 14 ICD-10 codes. 7 codes are represented by only one bacteria. The code A418 ("Other specified sepsis") contains 42 bacteria and A415 ("Other Gram-negative sepsis") contains 47.

A code has been found for each the bacteria in the list extracted from the lab. Only 2 bacteria were in the category "wrongly assigned code" : Actinobaculum schaalii and Sutterella wadsworthensis. The algorithm has an error rate of 1.22 %.

Sepsis caused by Actinobaculum schaalii was coded A414 by the algorithm, as a sepsis caused by an anaerobic bacteria. The Gold Standard says A418, which describes sepsis caused by gram-positive and not strictly anaerobic bacteria. No information about the aerobic status of this bacteria was found in SNOMED-CT. Sepsis caused by Sutterella wadsworthensis was coded A415 by the algorithm (gram-negative and aerobic bacteria), instead of A414 in the Gold Standard. No information about the respiratory status either anaerobic ou aerobic was found in SNOMED-CT.

Assignation mistakes are due to the absence of essential characeristics in SNOMED-CT for the classification. Furthermore, the concept of anaerobia is not clear and hard to discriminate from aerobia in some cases. For instance, the bacteria "Suterella wadsworthensis" is identified as a micro-aerophilic [6], and thus a sepsis would have been coded A415, like the algorithm returned but was classed as an anaerobic bacteria by the two microbiologists.

Conclusion

We developed a method allowing the coding of a sepsis from to the bacteria found in the haemoculture. This will dispense the coder to search the category of the bacteria and mobilize microbiology knowledge. A study of the impact on the precision of the coding has to be lead to verify the interest of the method, and its eventual scalability in other domains, presenting other caterogies of structured data (drugs treating a disease, procedures to treat a traumatism, etc.).

References

[1] L. Ahmadian, M. van Engen-Verheul, F. Bakhshi-Raiez, N. Peek, R. Cornet, and N. F. de Keizer, The role of standardized data and terminological systems in computerized clinical decision support systems : Literature review and survey, *Int J Med Inform 80* (2011), 81-93

[2] MF. Chiang, JC. Hwang, AC. Yu, DS. Casper, JJ. Cimino, and J. Starren, Reliability of SNOMED-CT Coding by Three Physicians using Two Terminology Browsers, *Amia Symp Proc* (2006), 131-135

[3] F. Bashki-Raiez, NF. de Keizer, R. Cornet, M. Dorrepaal, D. Dongelmans, and MWM. Jaspers, A usability evaluation of a SNOMED-CT based compositional interface terminology for intensive care, *Int J Med Inform 81* (2012), 351-362

[4] C. Duclos, J. Nobécourt, GL. Cartolano, A. Ellini, A. Venot, An ontology of bacteria to help physicians to compare antibacterial spectra, *Amia Symp Proc* (2007), 196-200

[5] JB. Lamy, A. Venot, and C. Dulcos, PyMedTermino : an open-source generic API for advanced terminology services, *Digital Healthcare Empowering Europeans* (2015), 924-928

[6] HM. Wexler, D. Reeves, PH. Summanen, E. Molitoris, M. McTeague, J. Duncan, KH. Wilson, and SM. Finegold, Sutterella wadsworthensis gen. nov., sp. nov., bile-resistant microaerophilic Campylobacter gracilis-like clinical isolates, *Int J Syst Bacteriol.*(1996), 252-8

Address for correspondence

Iris Ternois, LIMICS, 74 rue Marcel Cachin, 93017 Bobigny, France – iris.ternois@gmail.com