Diagnostic Accuracy Study

Medicine®

OPEN

Identifying multiple myeloma patients using data from the French health insurance databases

Validation using a cancer registry

Aurore Palmaro, PhD[a,b,c,*], Martin Gauthier, MSc[d], Cécile Conte, MSc[a,b], Pascale Grosclaude, MD, PhD[b,e,f], Fabien Despas, PharmD, PhD[a,b,c], Maryse Lapeyre-Mestre, MD, PhD[a,b,c]

Abstract

This study aimed to assess the performance of several algorithms based on hospital diagnoses and the long-term diseases scheme to identify multiple myeloma patients.

Potential multiple myeloma patients in 2010 to 2013 were identified using the presence of hospital records with at least 1 main diagnosis code for multiple myeloma (*ICD-10* "C90"). Alternative algorithms also considered related and associated diagnoses, combination with long-term conditions, or at least 2 diagnoses. Incident patients were those with no previous "C90" codes in the past 24 or 12 months. The sensitivity, specificity, and positive and negative predictive values (PPVs and NPVs) were computed, using a French cancer registry for the corresponding area and period as the criterion standard.

Long-term conditions data extracted concerned 11,559 patients (21,846 for hospital data). The registry contained 125 cases of multiple myeloma. Sensitivity was 70% when using only main hospital diagnoses (specificity 100%, PPV 79%), 76% when also considering related diagnoses (specificity 100%, PPV 74%), and 90% with associated diagnoses included (100% specificity, 64% PPV).

In relation with their good performance, selected algorithms can be used to study the benefit and risk of drugs in treated multiple myeloma patients.

Abbreviations: CI = confidence interval, dx = diagnosis, ICD-9/10 = International Classification of Diseases, 9th/10th version, ICD-O-3 = International Classification of Diseases for Oncology, 3rd Edition, IQR = interquartile range, LTD = long-term condition scheme, MM = multiple myeloma, OMOP = observational medical outcomes partnership, PMSI = Programme de médicalisation des systèmes d' information (Program for the Medicalization of Information Systems), NPV = negative predictive value, PPV = positive predictive value, SNIIRAM = Système national d'information inter-régime de l'assurance maladie (National inter-scheme information system on health insurance), T2A = tarification à l'activité (activity-based diagnosis Related Groups payment system).

Keywords: algorithms, cancer registry, electronic health records, multiple myeloma, pharmacoepidemiology, sensitivity and specificity

1. Introduction

Multiple myeloma (MM) is the second most common hematological malignancy in France. [1,2] In the last 2 decades, transplantation approaches and new drug regimens based on immunomodulatory drugs or proteasome inhibitors have considerably improved the survival of these patients.[3] These patients are now essentially treated as outpatients. Hospital-based observational studies are then no more sufficient to study real-life practices and patients' outcomes (adherence, among others.).

In parallel, researchers have access to the large French health insurance databases, covering >98% of the French population. French health insurance databases are potentially a valuable source for studying multiple myeloma epidemiology, healthcare use, and clinical outcomes, as it is among the rare automated databases in which certain hospital-administered medications are identifiable on an individual level. Indeed, the SNIIRAM ("*Système national d'information inter-régime de l'assurance maladie*") gathers ambulatory and hospital data. Its potential for research is also in relation with its national coverage and the availability of details on long-term conditions.

To implement epidemiological or pharmacoepidemiological studies on these patients, the validity of the coding is of primary

Editor: Ching-Sheng Hsu.

This work has been supported by the National Research Agency (ANR: Agence Nationale de la Recherche) for the "investissement d'avenir" (ANR-11-PHUC-001, CAPTOR project) and by "La ligue contre le Cancer."

The authors declare no conflict of interest.

Supplemental Digital Content is available for this article.

[a] Medical and Clinical Pharmacology Unit, Toulouse University Hospital, [b] INSERM 1027, University of Toulouse, [c] CIC 1436, Toulouse University Hospital, [d] Department of Hematology, Toulouse University Hospital, [e] Tarn Cancer Registry, Albi, [f] French Network of Cancer Registries (FRANCIM), France.

[*] Correspondence: Aurore Palmaro, Medical and Clinical Pharmacology Unit, Toulouse University Hospital, Pharmacoepidemiology Research Unit, INSERM 1027, University of Toulouse, 37, allées Jules Guesde, 31000 Toulouse, France (e-mail: aurore.palmaro@univ-tlse3.fr).

Medicine (2017) 96:12(e6189)

Received: 4 July 2016 / Received in final form: 16 December 2016 / Accepted: 3 February 2017

http://dx.doi.org/10.1097/MD.0000000000006189

importance.[4–6] As algorithms' performance could be in many ways database-specific, there was a need to implement this validation study in French health insurance databases. A lot of previous validations were made with the *ICD-9* in databases in the United States and validation studies are lacking for European and Nordic databases, in which *ICD-10* is more frequent.[6] Although several studies have measured the validity of cancer cases ascertainment in France,[7–9] none focused on hematological diseases. Then, the validity of identification of multiple myeloma cases through these databases has not been previously established. This study aimed to assess the performance of several algorithms based on hospital diagnoses (PMSI, "*Programme de médicalisation des systèmes d' information*") and diagnoses from the long-term diseases (LTDs) scheme.

2. Material and methods

2.1. Setting and design

We conducted a population-based and retrospective validation study of MM case ascertainment through health insurance records, using the Tarn cancer registry as the reference standard. For the Tarn Cancer registry, the period of interest was 2010 to 2013 (all incident cases diagnosed during 2010–2013). In health insurance records, date of diagnosis is not available, and an "observation period" (12 or 24 months here) without any diagnosis of interest is required to discriminate incident ("new") from prevalent patients. When no diagnosis of interest is recorded for up to 12 or 24 months, and then a first diagnosis occurs after this period, the patient is considered as incident. Then, for hospital diagnoses and long-term conditions, a longer extraction period has to be used (2008–2013) to enable at least 24 months (2008–2010) or 12 months (2009–2010) of observation (and use of the first occurrence of MM diagnosis after this period as a proxy of diagnosis date).

2.2. The Tarn Cancer registry

The Tarn Cancer Registry collects cancer data related to inhabitants of the Tarn area (about 400,000), an administrative area located in the southwest of France. Case ascertainment in the registry is based on systematic data collection from different sources: long-term diseases according to the health insurance schemes, hospital data for all residents of the Tarn area (all hospital data for the Tarn, plus data from hospital and reference centers in surrounding regions outside Tarn area), oncology regional network, pathology laboratories, hematology and cytology laboratories, all relevant hospital departments in public hospital or private clinics, radiotherapy centers, office from specialized physicians, and electoral registers.[10] Diagnoses are coded according to the International Classification of Diseases for Oncology, 3rd Edition (*ICD-O-3*).[11] The registry contains demographic details and some clinical or testing results. It also includes date of diagnosis attributed according international guidelines.[12]

Data from the registry were obtained for patients with hematological malignancies (ICD-O-3 topography code C42). Confirmed cases of multiple myeloma were patients diagnosed in 2010 to 2013 with ICD-O code "9732/3" for multiple myeloma, "9731/3" (for plasmacytoma), and "9734/3" (extramedullary plasmacytoma) in the registry. Clinical data were extracted for descriptive purposes.

2.3. Data from hospitalization stays PMSI

Data from hospitalization stays (PMSI) for the corresponding area and period (2010–2013 plus 2008–2009 or 2009 only for the observation period) were also obtained. Hospital data are managed within a single case-mix database of the activity-based payment system, ("*tarification à l'activité*," *T2A*). Data provided came from medical, surgical, and obstetrics care (PMSI MCO). PMSI provides data on all claims paid by the national health insurance system (covering >98% of the French population[13]) to public and private hospitals. Main, related, and associated diagnosis codes are coded according to the 10th version of the international classification of the diseases (*ICD-10*).[13] The data extraction was realized for hospital episodes involving a main diagnosis for cancer (*ICD-10* "C" or "D") or chemotherapy ("Z51").

2.4. Diagnoses from "long-term conditions" scheme

Data from long-term conditions (LTDs) *(affections de longue durée - ALD)* for all patients with cancer (*ICD*-10 "C" or "D") were extracted for the period 2008 to 2013, to enable at least 24 months (2008–2010) or 12 months (2009–2010) for the observation period. LTD provision is dedicated to patients suffering from a chronic condition which requires long-term treatment or expensive drugs. Healthcare expenses in relation to these conditions are fully covered. The list is established by decree (30 conditions), and include for instance malignant tumors, diabetes, or long-term psychiatric conditions. Diseases are coded according to *ICD*-10.[13] Entry in the LTD is obtained following a request by a physician (often the general practitioner) and is not systematically requested, in particular when the patient is already in the scheme for another disease. However, it is a common practice for researchers working on French healthcare databases to use LTD in combination with hospital diagnoses to improve sensitivity of disease identification or to measure comorbidities (Charlson score, among others).[13]

2.5. Data collected

Data from both data sources were obtained as nonanonymized data. Linkage between both sources was done on the basis of combinations of 5 potential matching variables: family name, birth name, first name, date of birth, sex, place of birth ("commune," lowest administrative area in France). Twenty-four possible combinations were tested. Unmatched patients were considered as having no hospital or LTD records during the period.

Nonanonymized hospital and LTD data have the same origin and structure as those contained in the national and anonymized health insurance database widely used for research (SNIIRAM). Combining hospital and LTD data at the local level is intended to simulate the performance of further algorithms that would be based on the SNIIRAM only.

2.6. Confidentiality and ethics

Data from hospitalization stays and long-term conditions were only those previously extracted for internal use of the Tarn Cancer Registry. All data were treated confidentially. The Tarn Cancer Registry is registered at the CNIL, French national data privacy institute (99 80 15 [12/1998], 99 80 15 version 2 [10/2003]).

Table 1

Characteristics of multiple myeloma patients in the Tarn cancer registry (*N* = 125).

	Multiple myeloma patients
Sex, n (%)	
Male	71 (56.8)
Female	54 (43.2)
Age, y	
median (IQR)	74 (63–81)
Durie-Salmon staging system, n (%)	
I	8 (6.3)
IA	8 (6.3)
IB	1 (0.8)
II	1 (0.8)
II	6 (4.8)
IIB	1 (0.8)
III	28 (22.2)
IIIA	28 (22.2)
IIIB	2 (1.6)
Missing	42 (33.6)
Myelogram performed, n (%)	117 (93.6)
Normal	8 (6.4)
Abnormal	105 (84.0)
Unknown	3 (2.4)
Karyotype performed, n (%)	36 (28.8)
Normal	15 (12.0)
Abnormal	10 (8.0)
Unknown	11 (8.8)

IQR = interquartile range.

2.7. Statistical analyses

Descriptive statistics were used to characterize the study population. Potential multiple myeloma patients in 2010 to 2013 were identified in the French health insurance databases using hospital records (PMSI) or LTD data. Indeed, the algorithm should use hospital data, but diagnoses are organized into 3 categories of diagnosis, and impacted by coding practices (main and related diagnoses are diagnoses of the current hospitalization, whereas associated diagnosis could be related to older episodes). When designing an algorithm, we have to include either main, main + related, or all types of diagnosis. As we did not have strong a priori on how these combinations will perform, we decided to test it systematically, for the 2 observation periods and with and without LTD to identify the combination providing the best performance. The algorithms tested began with a very straightforward approach (at least 1 main diagnosis), and then tested additional combinations and then cumulative diagnosis. Owing to the organization of the LTD scheme (long periods of coverage with start and end date), searching cumulative records was not relevant for this source.

In total, 13 algorithms corresponding to 3 strategies were tested with 2 different durations for defining incident patients (option A: 24-month observation period; option B: 12 months): Strategy 1, algorithms based on hospital data only (either main, main + related, or main + related + associated diagnoses), Strategy 2, algorithms based on hospital data or long-term condition (at least 1 long-term condition or either main, main + related, or main + related + associated diagnoses); and Strategy 3, cumulative diagnosis conditions for hospital data only (a 2nd MM main, main + related, or main + related + associated diagnosis, ≥30 days after the 1st). This cumulative condition was introduced with a temporal condition, as diagnoses belonging to the same episode (including transfers) are likely to be affected by the same potential coding error. True-positive patients were those ascertained as multiple myeloma cases in the registry (*ICD-O* code "9732/3," "9731/3," or "9734/3" in 2010–2013), and correctly identified as MM cases when applying the algorithm to health insurance data. True negative were those with no *ICD-O* code for MM in the registry, and not identified as MM cases according to the algorithm. False-positives were those not registered as MM cases in the registry (no corresponding ICD-O codes), but incorrectly identified as MM cases when using the algorithm based on health insurance data. False-negative were those ascertained as MM cases in the registry, but not identified as MM cases according to the algorithm. The sensitivity, specificity, positive predictive values (PPVs), and negative predictive values (PPVs) of the algorithms were then computed, using the cancer registry as the criterion standard. Exact binomial 95% confidence intervals were computed for each parameter. Youden index (sensitivity + specificity − 1) was computed as an indicator of model performance. Receiver-operating characteristic curves are provided as supplementary content, http://links.lww.com/MD/B612. Considering further use in pharmacoepidemiological research, specificity was prioritized over sensitivity to reduce the potential impact of misclassification on risk estimates.[14] Statistical analyses were performed using SAS 9.4 (SAS Institute Inc, Cary, NC). Method of validation was reported in accordance with the modified Standards for Reporting of Diagnostic Accuracy criteria.[15,16] Concordance between date of diagnosis in the registry and first "C90"multiple myeloma code in LTD or hospital database was assessed using median time between date of diagnosis in the registry and first multiple myeloma "C90" code in number of days.

3. Results

3.1. Patients

For the period 2010 to 2013, the registry contained 125 incident cases of multiple myeloma (including 7 cases coded as plasmacytoma). According to the characteristics presented Table 1, median age was 74 (interquartile range, 63–81) and 57% were male (n = 71). Half of the patients were classified as stage III or IIIa according to the Durie-Salmon system.

Long-term conditions data recorded for the corresponding area concerned 11,559 patients for the period 2008 to 2013. Hospital data were available for 21,846 inhabitants of the Tarn area (2008–2013). Data from the registry were obtained for 1069 patients. Computations were then made on the joint population of both data sources, (i.e., 22,083), as 1 patient could be in >1 data source). Among the 125 MM patients in the registry, 115 (92%) had at least 1 matching record in hospital data, and 68 (54%) had at least one match with LTD records.

Sensitivity, specificity, and predictive values are reported in Table 2 (option A: 24-month observation period without diagnosis to select incident patients) and Table 3 (option B: 12-month observation).

3.2. Algorithms performance using both data sources separately (strategy 1)

From 2010 to 2013, 112 patients were identified as incident cases using main diagnoses from PMSI data (128 when using main and

Table 2

Algorithms performance when using hospitalization or long-term conditions (option A: 24-month observation period*).

Case definition	True-positive/ false-positive	True-negative/ false-negative	Youden index	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)
Strategy 1							
Hospital data only							
≥1 MM main dx	88/24	21934/37	70.3	70.4 (62.4–78.4)	99.9 (99.8–99.9)	78.6 (70.1–86.2)	99.9 (99.8–99.9)
≥1 MM main or related dx	95/33	21925/30	75.8	76.0 (68.5–83.5)	99.8 (99.8–99.9)	74.2 (66.7–81.8)	99.9 (99.8–99.9)
≥1 MM main, related or associated dx	113/64	21894/12	90.1	90.4 (85.2–95.6)	99.7 (99.6–99.8)	63.8 (56.8–70.9)	99.9 (99.8–99.9)
Long-term conditions only							
≥1 MM dx	68/1	21957/57	54.4	54.4 (45.7–63.1)	100.0 (99.9–100.0)	98.6 (95.7–100.0)	99.7 (99.6–99.8)
Strategy 2							
Hospital + long-term conditions							
(≥1 MM main dx) OR ≥1 LTD MM dx	101/25	21933/24	80.7	80.8 (73.4–87.7)	99.9 (99.8–99.9)	80.2 (73.2–87.1)	99.9 (99.9–99.9)
(≥1 MM main or related dx) OR ≥1 LTD MM dx	103/34	21924/22	82.2	82.4 (75.7 – 89.1)	99.8 (99.8–99.9)	75.2 (68.0–82.4)	99.9 (99.8–99.9)
(≥ 1 MM main, related or associated dx) OR ≥1 LTD MM dx	113/64	21894/12	90.1	90.4 (85.2–95.6)	99.7 (99.6–99.8)	63.8 (56.8–70.9)	99.9 (99.9–99.9)
Strategy 3: cumulative diagnosis + period conditions							
Hospital data only							
2nd MM main dx, ≥30 days after the 1st	24/8	21950/101	19.1	19.2 (12.3–26.1)	99.9 (99.9–99.9)	75.0 (60.0–90.0)	99.5 (99.5–99.6)
2nd MM main or related dx, ≥30 days after the 1st	61/16	21942/64	48.7	48.8 (40.0–57.6)	99.9 (99.9–99.9)	79.2 (70.2–88.3)	99.7 (99.6–99.8)
2nd MM main, related or associated dx, ≥30 days after the 1st	77/29	21929/48	61.4	61.6 (53.1–70.1)	99.8 (99.7–99.9)	72.6 (64.2–81.1)	99.8 (99.7–99.9)
Hospital + long-term conditions							
(2nd MM main dx, ≥30 days after the 1st) OR ≥1 LTD MM dx	74/9	21949/51	59.2	59.2 (50.6–67.8)	100.0 (99.9–100.0)	89.2 (82.5–95.9)	99.8 (99.7–99.8)
(2nd MM main or related dx, ≥30 days after the 1st) OR ≥1 LTD MM dx	83/17	21941/42	66.3	66.4 (58.1–74.7)	99.9 (99.9–99.9)	83.0 (75.6–90.4)	99.8 (99.8–99.9)
(2nd MM main, related or associated dx, ≥30 days after the 1st) OR ≥1 LTD MM dx	89/30	21928/36	71.0	71.2 (63.3–79.1)	99.8 (99.8–99.9)	74.8 (67.0–82.6)	99.8 (99.7–99.9)

CI = confidence interval, dx = diagnosis, LTD = long-term condition scheme, MM = multiple myeloma, NPV = negative predictive value, PPV = positive predictive value.

* With no main, related, or associated dx in the past 24 months.

Table 3

Algorithms performance when using hospitalization or long-term conditions (option B: 12-month observation period*).

Case definition	True-positive/ false-positive	True-negative/ false-negative	Youden index	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)
Strategy 1							
Hospital data only							
≥1 MM main dx	88/25	21933/37	70.3	70.4 (62.4–78.4)	99.9 (99.8–99.9)	77.9 (70.2–85.5)	99.8 (99.7–99.8)
≥1 MM main or related dx	95/41	21917/30	75.8	76.0 (68.5–83.5)	99.8 (99.8–99.9)	69.9 (62.1–77.6)	99.8 (99.8–99.9)
≥1 MM main, related or associated dx	113/77	21881/12	90.1	90.4 (85.2–95.6)	99.7 (99.6–99.8)	59.5 (52.5–66.5)	99.9 (99.8–99.9)
Long-term conditions only							
≥1 MM dx	68/1	21957/57	54.4	54.4 (45.7– 63.1)	100.0 (99.9–100.0)	98.6 (95.7–100.0)	99.7 (99.6–99.8)
Strategy 2							
Hospital + long-term conditions							
(≥ 1 MM main dx) OR ≥1 LTD MM dx	101/26	21932/24	80.7	80.8 (73.9–87.7)	99.9 (99.8–99.9)	79.5 (72.5–86.6)	99.8 (99.8–99.9)
(≥1 MM main or related dx) OR ≥1 LTD MM dx	103/42	21916/22	82.2	82.4 (75.7–89.1)	99.8 (99.7–99.9)	71.0 (63.7–78.4)	99.9 (99.8–99.9)
(≥ 1 MM main, related or associated dx) OR ≥1 LTD MM dx	113/77	21881/12	90.1	90.4 (85.2–95.6)	99.7 (99.6–99.7)	59.5 (52.5–66.5)	99.9 (99.9–100.0)
Strategy 3: cumulative diagnosis + period conditions							
Hospital data only							
2nd MM main dx, ≥30 days after the 1st	24/8	21950/101	19.1	19.2 (12.3–26.1)	99.9 (99.9–99.9)	75.0 (60.0–90.0)	99.5 (99.4–99.6)
2nd MM main or related dx, ≥30 days after the 1st	61/24	21934/64	48.7	48.8 (40.0–57.6)	99.9 (99.9–99.9)	71.8 (62.2–81.3)	99.7 (99.6–99.8)
2nd MM main, related or associated dx, ≥30 days after the 1st	77/39	21919/48	61.4	61.6 (53.1–70.1)	99.8 (99.8–99.9)	66.4 (57.8–75.0)	99.8 (99.7–99.9)
Hospital + Long-term conditions							
(2nd MM main dx, ≥30 days after the 1st) OR ≥1 LTD MM dx	74/9	21949/51	59.1	59.2 (50.6–67.8)	99.9 (99.9–99.9)	89.2 (82.5–95.9)	99.6 (99.5– 99.6)
(2nd MM main or related dx, ≥30 days after the 1st) OR ≥1 LTD MM dx	83/25	21933/42	66.3	66.4 (58.1–74.7)	99.9 (99.8–99.9)	76.9 (68.9–84.8)	99.7 (99.6–99.8)
(2nd MM main, related or associated dx, ≥30 days after the 1st) OR ≥1 LTD MM dx	89/40	21918/36	71.0	71.2 (63.3–79.1)	99.8 (99.8–99.9)	69.0 (61.0–77.0)	99.8 (99.7–99.8)

CI = confidence interval, dx = diagnosis, LTD = long-term condition scheme, MM = multiple myeloma, NPV = negative predictive value, PPV = positive predictive value.

* With no main, related, or associated dx in the past 12 months.

Table 4

Concordance between date of diagnosis in the registry and first multiple myeloma "C90" code in healthcare database.

Median duration in days (IQR)	12 mo	24 mo
Delays between first hospital diagnosis and date of diagnosis in the registry		
Main dx	0 (−2; 55)	0 (−2; 65)
Main or related dx	0 (−2; 51)	0 (−2; 39)
Main, related or associated dx	0 (−2; 49)	0 (−3; 21)
Delays between first long-term condition and date of diagnosis in the registry	−4 (−16; 6)	−4 (−16; 7)

IQR = interquartile range.

related diagnoses, and 177 with associated diagnoses) when using a 24-month observation period (option A, Table 2). Sensitivity was 70.4% (62.4%–78.4%) when using only main hospital diagnoses (specificity 99.9%, PPV 78.6%), 76.0% (68.5%–83.5%) when considering also related diagnoses (specificity 99.9%, PPV 74.2%), and 90.4% (85.2%–95.6%) with associated diagnoses included (99.7% specificity, 63.8% PPV).

Using a 12-month observation period (option B, Table 3) gave very close results, with similar sensitivity (90.4%) and slight differences visible for the PPV (59.5% vs. 63.8%), but with overlapping confidence intervals.

3.3. Impact of long-term conditions (strategy 2)

LTD alone exhibited very poor performance, with sensitivity around 55% whatever the period of observation used. In the algorithm considering a 24-month period to define incident cases (option A, Table 2), sensitivity was increased by up to 10% when incorporating long-term conditions to main hospital diagnoses. However, the interest of long-term conditions was attenuated after integrating associated diagnoses (+6%), and disappeared in all algorithms integrating associated diagnoses (same value for sensitivity for algorithm with main, related or associated code [90%], with or without LTD, 12 or 24 months' period).

Using a 12-month observation period did not impact the performance of the algorithm (option B, Table 3).

3.4. Impact of the number of diagnoses required (strategy 3)

When a second diagnosis was required >30 days after a first diagnosis (24-month observation period; option A, Table 2), sensitivity dropped dramatically to very low values (19%). Impact of specificity was not observable as it was already maximal (99%) for the algorithm with only one diagnosis required. Among the 88 true-positive patients identified using the first algorithm (at least 1 main hospital diagnosis, 24-month period), 62% (55/88) have only 1 hospital diagnosis and 38% (33/88) ≥1 main hospital diagnoses.

The performance of this strategy for a 12-month observation period (option B, Table 3) was similar.

3.5. Impact of period of observation (24 vs. 12-month observation period)

There was no decrease in sensitivity and very slight reductions in PPV (<3%) as the observation window increased from 12 to 24 months (Table 3 vs. Table 2). Varying the window between MM codes and exclusion criteria did not improve algorithm performance. The algorithm using a 12-month observation period (option B) and "at least 1 main, OR related, OR associated hospital MM code" (strategy 1) exhibited the same highest performance (Youden's index: 90.1) as compared to the same algorithm with a 24-month period (Tables 2 and 3).

3.6. Exploration of diagnoses in false incidents

Using the first algorithm (at least 1 main diagnosis, option A: 24 months), 24 patients were classified as false-positive cases (patients misclassified as having MM). Among these false-positives, 2 were identified in the registry, with *ICD-O-3* diagnoses of plasmablastic lymphoma ("9735/3") and refractory thrombocytopenia ("9992/3"). All other false-positive patients were patients with hospital data, but not retrieved in the registry. When looking at their other hospital diagnoses, no other code for distinct hematological malignancies was retrieved. Two patients had diagnoses for bone metastasis (*ICD-10* "C79").

For the false-negative patients for algorithms using only main diagnoses (respectively, main or related), all appeared to be incident MM cases that would be selected when using either or related or associated hospital code (respectively, associated codes). Finally, almost all false-negatives had no available hospital records, and only 1 ascertained MM case had another hospital diagnosis (D46.2: refractory anemia with excess of blasts).

3.7. Exploration of delays between first hospital record and date of diagnosis in the registry

When considering time between diagnosis in the registry and date of first MM diagnosis in hospital data (Table 4), correspondence was high, with a median time of 0 days (interquartile range -2; 21 for main, related or associated diagnoses and a 24-month period), and 94% (83/88) of patients with a main diagnosis between 30 days before and within 1 year after registry documented MM diagnosis (72% between 30 days before and 30 days after, 63/88).

4. Discussion

4.1. Main findings

Algorithms tested exhibited very different performances, ranging from poor performance when using only main hospital diagnoses to very acceptable parameters when hospital data are used in combination with long-term conditions diagnoses. The optimal algorithm to identify MM patients (maximizing both Youden index and specificity) was "at least 1 main, OR related, OR associated hospital MM code," with a 12-month observation period, which had a sensitivity of 90%, a specificity of 100%, and a PPV of 60%. The same algorithm with a 24-month observation period demonstrated similar performance, but the algorithm with

the shorter period of observation should be preferred. In the same way, one of the algorithms for strategy 2 performed equally well ("at least 1 MM main, related or associated hospital MM code) OR at least 1 LTD MM code"), but would require to have LTD data available. Faced to these 2 algorithms with equal performances, we choose the one requiring the minimum data (i.e., with no participation of data from the LTD scheme).

Indeed, the study design simulated the performance of algorithms that would be based on the large French health insurance databases (SNIIRAM) in further research. Using an algorithm with a restricted period of observation (12 months as compared to 24 months) has potentially a great interest for increasing sample size and length of possible follow-up in the context of limited longitudinal data availability (data are available since 2006 in the SNIIRAM).

4.2. Strengths and limitations

Some limitations must be acknowledged. First, this study was conducted in a single area and may not perfectly reflect the performance of hospital and LTD codes at the national level. However, coding is quite similar in all private and public hospitals. However, several data suggest that clinical and coding practices in Tarn region are very close to that of the whole nation. According to national estimations, 51% of newly diagnosed MM (or proliferative disorders) were male, with 70% aged >65 years. [1] In our study, 58% were male and 67% were older than 65 years. Sex and gender characteristics are then quite close to the national reference. Clinical features of the patients are not expected to be different in France, but might be acknowledged at the larger scale (i.e., at the European level) owing to different delay to diagnosis. As we only used specific MM codes and not symptoms, the performance of our algorithm is not likely to be affected by the clinical aspect. Concerning coding practices for health insurance data, there are national standards of coding PMSI data, for the choice of the principal diagnosis for instance. Of course, we could not rule out various coding habits between hospitals or regions because of different interpretation of coding rules. This study is then conducted with the assumption of similar coding practices in all hospitals. However, in practice, quality of coding is regularly audited in public and private hospitals for reimbursement purposes.

Another potential limitation is related to the assumption that all patients with available data from the registry corresponding area and period should be considered as potential cases. In other words, failure of linkage with administrative data meant that the patient had no LTD or hospital record for the period of interest. However, even if there is systematic attempt to obtain cancer hospital data from all inhabitants of the covered area (Tarn), including hospitals outside the area of the residence, there is a possibility that hospital episodes were not complete, leading to underestimate the performance of the algorithm (increase in false-negative). Differences were observed when matching records from the cancer registry and the 2 data sources (21,846 for hospital data vs. 11,559 for LTD). These variations were expected. Indeed, although coding of hospital episodes will be always performed, LTD status is not automatically assigned to a patient and has to be requested by a physician. Then, the greater number of hospital records compared to LTD records should not be interpreted as missing information, but simply reflect the organization of the healthcare system.

In addition, the relative lack of sensitivity of the algorithm was expected owing to the particular natural history of multiple myeloma. Indeed, patients are not systematically managed in hospital, nor treated after diagnosis, because of possible asymptomatic or smoldering disease. Those patients are less likely to go to hospital or to enter in a long-term condition scheme. Lack of sensitivity for identifying these patients could be problematic for epidemiological purposes, but is acceptable for healthcare use or pharmacoepidemiological research, as these patients are not healthcare users.

In this study, we focused on the need to accurately identify myeloma patients (specificity), which would undergo additional selection criteria to be included in pharmacoepidemiological study for instance. Algorithms integrating treated patients would certainly have been more sensitive[17], but we did not have sufficient information in the registry to ascertain this. Finally, lack of sensitivity is likely to be controlled in additional selection process, and the high specificity of the algorithm is an important strength here.

The false-positives for non-MM patients having at least 1 MM diagnosis code may be related to testing for disease rather than confirmed disease or to coding errors, or perhaps may be because of the evolution of diagnoses over time. There would be a potential interest to require not only one but several diagnosis codes ("cumulative diagnoses" strategy) to overcome this issue. Then, as expected, the increase in the number of diagnosis codes further improved specificity, but sensitivity reached unacceptable levels (<60%). In practice, MM patients may not be hospitalized several times after initial diagnosis (62% of confirmed MM cases has only 1 hospital diagnosis in our study), thus limiting the relevance of algorithms requiring multiple diagnosis codes.

Another consideration in relation to hematological malignancies is the possibility of lack of recording in the registry, as shown for myeloid malignancy in the United States.[18,19]

4.3. Other French experiences

Data from the French health insurance database have already been used for epidemiological identification of cancer,[20] including for instance breast,[21] colorectal,[9] prostate,[22] thyroid,[7] or central nervous system malignancies.[8] A work has also been implemented in French hospital data to select cancer related hospitalization, and myeloma was listed among the diseases of interest.[23]

4.4. International experiences

Validation of case identification algorithms represents an important issue, as demonstrated by recent calls,[6] and also by several initiatives from Mini Sentinel and OMOP (Observational Medical Outcomes Partnership) in United States or EU-ADR in Europe.[24] An important series of systematic review on methods for validating a wide range of disease, including lymphoma for instance,[25] has been published since 2012.[26–32] Lessons learned and proposal for improvement have been formulated during these validation studies.[33] However, literature concerning multiple myeloma is very poor, and only one resource could be identified.[34] According to this study, on MarketScan databases linked to Medicare claims databases, at least 2 diagnoses provided a sensitivity of 95%, a specificity of 73%, and a positive predictive value (PPV) of 76%.

5. Conclusions

This study revealed that including simultaneously main, related, and associated hospital diagnoses increased the sensitivity of the

algorithm without generating excess false positives. The optimal algorithm to identify MM patients was "at least 1 main, OR related, OR associated hospital MM code," with a 12-month observation period, which had a sensitivity of 90%, a specificity of 100%, and a PPV of 60%. This algorithm can be used in further pharmacoepidemiological studies for investigating the benefits and risks of drugs used by multiple myeloma patients.

References

[1] INVSIncidence et survie des hémopathies malignes: données générales et situation chez les plus de 75 ans, France, 1989-1997. Bull Epidemiol Hebd 2007;09–10. 65-83.

[2] INVS. Projections de l'incidence et de la mortalité par cancer en France en 2011. Myélome multiple et maladie immunoprolifération. Available at: http://www.invs.sante.fr/applications/cancers/projections2010/donnees_localisation/myelome_multiple.pdf. Accessed June 23, 2016.

[3] Kumar SK, Rajkumar SV, Dispenzieri A, et al. Improved survival in multiple myeloma and the impact of novel therapies. Blood 2008;111:2516–20.

[4] Schulman KL, Berenson K, Shih Y-C, et al. A checklist for ascertaining study cohorts in oncology health services research using secondary data: report of the ISPOR Oncology Good Outcomes Research Practices Working Group. Value Health 2013;16:655–69.

[5] Manuel DG, Rosella LC, Stukel TA. Importance of accurately identifying disease in studies using electronic health records. BMJ 2010;341:c4226.

[6] Ehrenstein V, Petersen I, Smeeth L, et al. Helping everyone do better: a call for validation studies of routinely recorded health data. Clin Epidemiol 2016;8:49–51.

[7] Hafdi-Nejjari Z, Couris C-M, Schott A-M, et al. Role of hospital claims databases from care units for estimating thyroid cancer incidence in the Rhône-Alpes region of France. Rev Epidemiol Sante Pub 2006;54:391–8.

[8] Coureau G, Baldi I, Savès M, et al. Performance evaluation of hospital claims database for the identification of incident central nervous system tumors compared with a cancer registry in Gironde, France, 2004. Rev Epidemiol Sante Pub 2012;60:295–304.

[9] Quantin C, Benzenine E, Hagi M, et al. Estimation of national colorectal-cancer incidence using claims databases. J Cancer Epidemiol 2012;2012: e298369.

[10] Registre des cancers du Tarn. Health Databases. Available at: https://epidemiologie-france.aviesan.fr/epidemiologie-france/fiches/tarn-cancer-registry-certified-registry-2010-2013. Accessed June 23, 2016.

[11] Percy C, Fritz A, Jack A, et al. International Classification of Diseases for Oncology. 3rd ed.World Health Organization, Geneva:2000.

[12] Tyczynski JE, Demaret E, Parkin DM. Standards and guidelines for cancer registration in Europe. The ENCR recommendations. Volume 1. IARC Technical Publication No 40. International Agency for Research on Cancer, Lyon, France: 2003.

[13] Moulis G, Lapeyre-Mestre M, Palmaro A, et al. French health insurance databases: what interest for medical research? Rev Med Interne 2014;36:411–7.

[14] Chubak J, Pocobelli G, Weiss NS. Trade-offs between accuracy measures for electronic healthcare data algorithms. J Clin Epidemiol 2012;65: 343–9.

[15] Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Ann Intern Med 2003;138:W1–2.

[16] Benchimol EI, Manuel DG, To T, et al. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. J Clin Epidemiol 2011;64:821–9.

[17] Teitelbaum A, Ba-Mancini A, Huang H, et al. Health care costs and resource utilization, including patient burden, associated with novel-agent-based treatment versus other therapies for multiple myeloma: findings using real-world claims data. Oncologist 2013; 18:37–45.

[18] Craig BM, Rollison DE, List AF, et al. Underreporting of myeloid malignancies by United States cancer registries. Cancer Epidemiol Biomarkers Prev 2012;21:474–81.

[19] Cogle CR, Craig BM, Rollison DE, et al. Incidence of the myelodysplastic syndromes using a novel claims-based algorithm: high number of uncaptured cases by cancer registries. Blood 2011;117:7121–5.

[20] Grosclaude P, Dentan C, Trétarre B. Relevance of health administrative databases in cancer surveillance. Comparison with registries records at individual level. Bull Epidemiol Hebd 2012;63–7.

[21] Remontet L, Mitton N, Couris CM, et al. Is it possible to estimate the incidence of breast cancer from medico-administrative databases? Eur J Epidemiol 2008;23:681–8.

[22] Couris CM, Seigneurin A, Bouzbid S, et al. French claims data as a source of information to describe cancer incidence: predictive values of two identification methods of incident prostate cancers. J Med Syst 2006;30:459–63.

[23] Institut National Du Cancer. Algorithme de sélection des hospitalisations liées à la prise en charge du cancer dans les bases nationales d'activité hospitalière de court séjour. Available at: http://www.e-cancer.fr/Expertises-et-publications/Catalogue-des-publications/Algorithme-de-selection-des-hospitalisations-liees-a-la-prise-en-charge-du-cancer-dans-les-bases-nationales-d-activite-hospitaliere-de-court-sejour. Accessed June 23, 2016.

[24] Avillach P, Coloma PM, Gini R, et al. Harmonization process for the identification of medical events in eight European healthcare databases: the experience from the EU-ADR project. J Am Med Inform Assoc 2013;20:184–92.

[25] Herman RA, Gilchrist B, Link BK, et al. A systematic review of validated methods for identifying lymphoma using administrative data. Pharmacoepidemiol Drug Saf 2012;21(suppl 1):203–12.

[26] Andrade SE, Harrold LR, Tjia J, et al. A systematic review of validated methods for identifying cerebrovascular accident or transient ischemic attack using administrative data. Pharmacoepidemiol Drug Saf 2012;21 (suppl 1):100–28.

[27] Schneider G, Kachroo S, Jones N, et al. A systematic review of validated methods for identifying erythema multiforme major/minor/not otherwise specified, Stevens-Johnson Syndrome, or toxic epidermal necrolysis using administrative and claims data. Pharmacoepidemiol Drug Saf 2012;21 (suppl 1):236–9.

[28] Schneider G, Kachroo S, Jones N, et al. A systematic review of validated methods for identifying hypersensitivity reactions other than anaphylaxis (fever, rash, and lymphadenopathy), using administrative and claims data. Pharmacoepidemiol Drug Saf 2012;21(suppl 1):248–55.

[29] Carnahan RM, Moores KG, Perencevich EN. A systematic review of validated methods for identifying infection related to blood products, tissue grafts, or organ transplants using administrative data. Pharmacoepidemiol Drug Saf 2012;21(suppl 1):213–21.

[30] Kee VR, Gilchrist B, Granner MA, et al. A systematic review of validated methods for identifying seizures, convulsions, or epilepsy using administrative and claims data. Pharmacoepidemiol Drug Saf 2012;21 (suppl 1):183–93.

[31] Walkup JT, Townsend L, Crystal S, et al. A systematic review of validated methods for identifying suicide or suicidal ideation using administrative or claims data. Pharmacoepidemiol Drug Saf 2012;21 (Suppl 1):174–82.

[32] Williams SE, Carnahan R, Krishnaswami S, et al. A systematic review of validated methods for identifying transverse myelitis using administrative or claims data. Vaccine 2013;31(suppl 10):K83–7.

[33] Carnahan RM, Moores KG. Mini-Sentinel's systematic reviews of validated methods for identifying health outcomes using administrative and claims data: methods and lessons learned. Pharmacoepidemiol Drug Saf 2012;21(suppl 1):82–9.

[34] Princic N, Chris G, Willson T, et al. Development of an Algorithm to Identify Patients with Multiple Myeloma Using Administrative Claims Data. ASH Orlando 2015, 57th American Society of Hematology Annual Meeting, United States of America, Orlando, 5-8 December 2015. Available at: https://ash.confex.com/ash/2015/webprogram/Paper85570.html. Accessed June 23, 2016.