



## **Using healthcare claims data to analyze the prevalence of BCR-ABL-positive chronic myeloid leukemia in France: A nationwide population-based study**

Stéphanie Foulon, Pascale Cony-Makhoul, Agnès Guerci-Bresler, Marc Delord, Eric Solary, Alain Monnereau, Julia Bonastre, Pascale Tubert-Bitter

### **► To cite this version:**

Stéphanie Foulon, Pascale Cony-Makhoul, Agnès Guerci-Bresler, Marc Delord, Eric Solary, et al.. Using healthcare claims data to analyze the prevalence of BCR-ABL-positive chronic myeloid leukemia in France: A nationwide population-based study. *Cancer Medicine*, 2019, 8 (6), pp.3296-3304. <10.1002/cam4.2200>. <hal-03212260>

**HAL Id: hal-03212260**

**<https://hal.science/hal-03212260v1>**

Submitted on 29 Apr 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.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

# Using healthcare claims data to analyze the prevalence of BCR-ABL-positive chronic myeloid leukemia in France: A nationwide population-based study

Stéphanie Foulon<sup>1,2,3</sup>  | Pascale Cony-Makhoul<sup>4,5</sup> | Agnès Guerci-Bresler<sup>5,6</sup> | Marc Delord<sup>5,7</sup>  | Eric Solary<sup>8,9</sup> | Alain Monnereau<sup>10,11,12</sup> | Julia Bonastre<sup>1,2</sup> | Pascale Tubert-Bitter<sup>3</sup>

<sup>1</sup>Biostatistics Unit, Gustave Roussy, Villejuif, France

<sup>2</sup>CESP Centre for Research in Epidemiology and Population Health, INSERM U1018, Paris-Sud Univ, Villejuif, France

<sup>3</sup>B2PHI Biostatistics, Biomathematics, Pharmacoepidemiology and Infectious Diseases, Inserm U1181, UVSQ, Paris Saclay Univ, Villejuif, France

<sup>4</sup>Service d'Hématologie, CH Annecy Genevois, Pringy, France

<sup>5</sup>FiLMC Group, Institut Bergonié, Bordeaux, France

<sup>6</sup>Service d'Hématologie, CHRU Brabois, Vandoeuvre, France

<sup>7</sup>Institut Universitaire d'Hématologie, Université Paris-Diderot Paris7, Paris, France

<sup>8</sup>Department of Hematology, Gustave Roussy, Villejuif, France

<sup>9</sup>INSERM U1170, Villejuif, France

<sup>10</sup>University of Bordeaux, Inserm, Bordeaux Population Health Research Center, Team EPICENE, UMR 1219, Bordeaux, France

<sup>11</sup>Registre des Hémopathies Malignes de la Gironde, Institut Bergonié, Bordeaux, France

<sup>12</sup>French Network of Population-based Cancer Registries (FRANCIM), Toulouse, France

## Correspondence

Stéphanie Foulon, Service de biostatistique et d'épidémiologie, Gustave Roussy, 114 rue Edouard Vaillant 94805 Villejuif Cedex, France.

Email: stephanie.foulon@gustaveroussy.fr

## Funding information

This study was funded by the French National Agency for Research (ANR-10-IBHU-0001) and the French Institute for Public Health Research.

## Abstract

**Background:** Data on Chronic Myeloid Leukemia (CML) prevalence are scarce. Here we provide an estimation of the prevalence of CML in France for the year 2014 using French national health insurance data.

**Methods:** We selected patients claiming reimbursement for tyrosine kinase inhibitors (TKI) or with hospital discharge diagnoses for CML, BCR/ABL-positive or with full reimbursement of health care expenses for myeloid leukemia. We built an algorithm which we validated on a random sample of 100 potential CML patients by comparing the results obtained using the algorithm and the opinion of two hematologists who reviewed the patient demographics and sequence of care abstracted from claims data (internal validity). For external validity, we compared the number of incident CML patients identified using the algorithm with those recorded in French population-based cancer registries in departments covered by such a registry.

Julia Bonastre and Pascale Tubert-Bitter contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Cancer Medicine* published by John Wiley & Sons Ltd.

**Results:** We identified 10 789 prevalent CML patients in 2014, corresponding to a crude prevalence rate of 16.3 per 100 000 inhabitants [95% confidence interval (CI) 16.0-16.6]; 18.5 in men [18.0-19.0] and 14.2 in women [13.8-14.6]. The crude CML prevalence was less than 1.6 per 100 000 [1.2-2.0] under age 20, increasing to a maximum of 48.2 [45.4-51.2] at ages 75-79. It varied from 10.2 to 23.8 per 100 000 across French departments. The algorithm showed high internal and external validity. Concordance rate between the algorithm and the hematologists was 96%, and the numbers of incident CML patients identified using the algorithm and the registries were 162 and 150, respectively.

**Conclusion:** We built and validated an algorithm to identify CML patients in administrative healthcare databases. In addition to prevalence estimation, the algorithm could be used for future economic evaluations or pharmaco-epidemiological studies in this population.

#### KEYWORDS

cancer registries, chronic myeloid leukemia, epidemiology, insurance claims database, prevalence

## 1 | INTRODUCTION

Chronic Myeloid Leukemia (CML) is a myeloproliferative neoplasm whose age-adjusted annual incidence rate is around 1 per 100 000 person-years in European countries.<sup>1,2</sup> Since the introduction of tyrosine kinase inhibitors (TKI) in the early 2000s, the survival of patients with CML has improved dramatically.<sup>3-7</sup> As CML incidence increases with age, an aging population combined with an improvement in survival is expected to increase CML prevalence. Based on the estimations from 22 cancer registries in Europe, the prevalence of CML was estimated to be 5.6 per 100 000 in 2008.<sup>8</sup> In Sweden, CML prevalence was estimated at 5.7 per 100 000 in 2000 and 11 per 100 000 in 2012.<sup>4,9</sup> In the United States, the number of CML cases was estimated to be 70 000 in 2010, indicating a high prevalence of 22.6 per 100 000 inhabitants, with a predicted number of cases reaching 112 000 in 2020.<sup>10</sup> In France, a regional hospital-based evaluation of CML prevalence (Nord-Pas de Calais region) indicated an increase in CML prevalence from 5.8 to 10.4 per 100 000 inhabitants between 1998 and 2007.<sup>11</sup> In France, CML prevalence was recently modeled from 1960 to 2060 by using incidence rates from six French population-based cancer registries (corresponding to 11 departments and covering 14% of the French population). Scenarios combining projections of the French population and various hypotheses on the evolution of relative survival of CML patients were proposed. In the base case scenario, CML prevalence was estimated at 2.5 per 100 000 inhabitants before the 1980s, 6.4 in 2002, 13.7 in 2012, and 17.5 in 2018, with an anticipated plateau at 32 per 100 000 inhabitants in the 2060s.<sup>12</sup> These estimates of CML prevalence were obtained through a modeling exercise using data

on relative survival from other European countries. Modeling estimates of CML prevalence were consistent with the number of imatinib sales in 2004 and the regional estimate of CML prevalence in Nord-Pas de Calais until 2007. But since then, no recent data were available to validate modeling estimates.

Administrative healthcare databases are regularly used to assess the economic burden of diseases and to allocate resources both at the national and the regional level. More recently, these data sources have been extensively used in economic evaluation to estimate input parameters in cost-effectiveness studies. Similarly, administrative and claims data are used to estimate target populations and market shares for pricing and reimbursement of new medicines as well as postmarketing surveillance studies. In addition, administrative databases are relevant sources of data to conduct epidemiological studies worldwide.<sup>13-18</sup> In the present study, our objective was to build and validate an algorithm to assess the prevalence of CML at the nationwide level, using individual data from French national insurance databases. Our aim was also to make available an algorithm to identify CML patients in administrative healthcare databases that could be used for future economic evaluations or pharmaco-epidemiological studies in this patient population.

## 2 | PATIENTS AND METHODS

### 2.1 | Study design and data sources

We performed a cross-sectional study using data from the French national health insurance database linked with the national hospital discharge database.<sup>19-21</sup> These databases

contain individualized, anonymous, and comprehensive data on all healthcare reimbursements (including hospital stays coded with the International Classification of Diseases, 10th Revision (ICD-10) diagnosis codes, medications, general practitioner and specialist consultations, imaging and biological procedures and sick/disability leaves) and patient eligibility for full reimbursement of health care expenses related to specific costly or long-term diseases (LTD) also coded with ICD-10 codes. Claims data are gathered across nearly all French Health Insurance schemes, covering 98.8% of the population living in France (66.4 million inhabitants on 1 January 2015). Besides healthcare reimbursement data, demographic data are also available including the year of birth, gender, area of residence, and date of death. There was no requirement for ethics approval to be sought for this observational study, based on anonymous healthcare claims data.

## 2.2 | Algorithm to identify CML patients

### 2.2.1 | Selection of potential CML patients

For the 2006–2014 period, we selected from the French national health insurance databases: (a) all patients treated by TKI indicated in CML (imatinib, dasatinib, nilotinib, bosutinib or ponatinib) and/or (b) identified by the ICD-10 diagnosis code C92.1 (Chronic Myeloid Leukemia, BCR/ABL-positive) among hospital discharge diagnoses and/or (c) having requested full reimbursement of health care expenses for Myeloid Leukemia (ICD-10 diagnosis code C92). The latter criterion (ICD-10 diagnosis code C92 for Myeloid Leukemia) is coded only with three characters in the French national health insurance databases. It is therefore not specific for Chronic Myeloid Leukemia and patients with a very different disease such as Acute Myeloid Leukemia could also be selected using this ICD-10 code. However, at the time of the data extraction, we preferred not to be too restrictive in our selection criteria. Then, to refine this selection, we examined the sequence of care between 2006 and 2014 for 200 randomly selected individuals, using a standardized form (Figure S1). For each individual, this form presented demographic data and healthcare resource utilization.

### 2.2.2 | Algorithm building

Based on this thorough examination and clinical opinions, we built a claim-based algorithm to identify CML patients. Case definition was based on (a) identifying any TKI reimbursement associated with a hospitalization for CML or lasting  $\geq 2$  months and (b) excluding patients receiving TKIs for diseases other than CML including Philadelphia positive Acute Lymphoblastic Leukemia, Gastrointestinal Stromal Tumor, Graft versus Host Disease, Hyper Eosinophilic Syndrome and Stromal or other Connective Tissue Tumor. These differential

diagnoses were identified using discharge diagnoses, Long-Term Disease diagnoses and reimbursement of specific drugs (Figure S2). The SAS program of the algorithm is available upon request from the corresponding author.

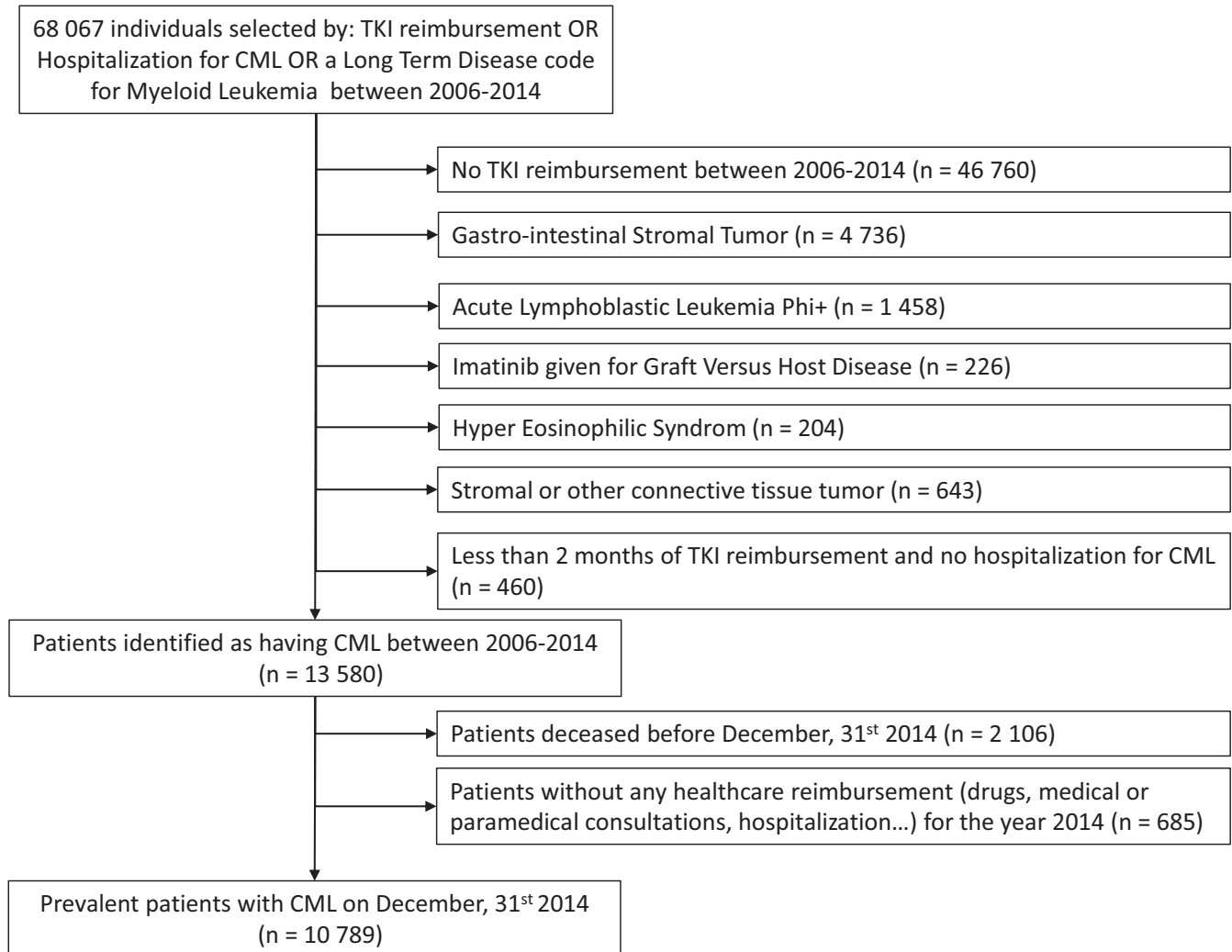
## 2.3 | CML prevalence

Prevalent CML cases were those identified using the algorithm described above, having at least one healthcare reimbursement during the year 2014 and still alive on 31 December 2014. The nationwide crude CML prevalence rate was defined as the number of prevalent CML patients divided by the number of people living in France on 1 January 2015 ([www.insee.fr](http://www.insee.fr))<sup>22</sup> and is provided with an exact Clopper-Pearson confidence interval.<sup>23</sup> The national CML prevalence rate was also standardized by gender and by 5-year age group to the 1976 European population, and the 1960 World population, to allow for international comparisons.<sup>24</sup> Both crude and standardized rates correspond to estimates of the complete prevalence of CML. Crude and standardized prevalence rates by department, using the department of residence of prevalent CML patients in 2014, were computed to explore geographical variations in CML prevalence in the French territory. The reference population for the standardization on gender and age was the general population residing in France on 1 January 2015.

## 2.4 | Algorithm validity

The internal validity of our algorithm was assessed from another random sample of 100 potential CML patients. We calculated the proportion of agreement between the result of the algorithm and the opinion of two expert hematologists from the FiLMC group. For each individual, hematologists reviewed the standardized patient form presenting patient demographics and sequence of care from claims data.

In addition, we performed an external validation of our algorithm using data from French cancer registries. We compared the number of incident CML cases (ie new CML cases) identified from the French national health insurance databases for 3 years from 2012 to 2014 with the number of incident CML cases (ICD-O-3 9875/3: Chronic myelogenous leukemia, BCR/ABL-positive) recorded in 18 French population-based cancer registries for the same years and the same geographical areas (covering 22% of the French population). In the French national health insurance databases, the index date for the diagnosis of CML was defined as the first date among the first date of TKI or hydroxycarbamide reimbursement, or the start date for full reimbursement of health care expenses for myeloid leukemia or the first date of hospitalization for CML. In a sensitivity analysis, the CML prevalence rate was adjusted using the results of external validation (cf. Supplementary). Statistical analyses were performed using SAS 9.3.



**FIGURE 1** Patient flow chart. Data were extracted from the French national health insurance databases for the period 2006-2014. TKI: Tyrosine Kinase Inhibitor, CML: Chronic Myeloid Leukemia

### 3 | RESULTS

#### 3.1 | Population study

68 067 individuals were selected from the French national health insurance databases using the selection criteria described in the methods section (paragraph 2.2.1). Applying the algorithm (paragraph 2.2.2) resulted in the identification of 10 789 patients with CML on 31 December 2014 (Figure 1). Forty-eight percent of these prevalent CML cases had the three selection criteria (TKI, hospitalization with a CML diagnosis code, full reimbursement of health care expenses for myeloid leukemia), 32% were treated by TKI and benefited from full reimbursement of health care expenses for myeloid leukemia, 9% had TKI and hospitalization with CML diagnosis code. Finally, 11% of the CML patients identified using the algorithm were captured by TKI reimbursement only. The median age [interquartile range] of the prevalent population of CML patients was 63 [51-73], with a slight male preponderance (55%).

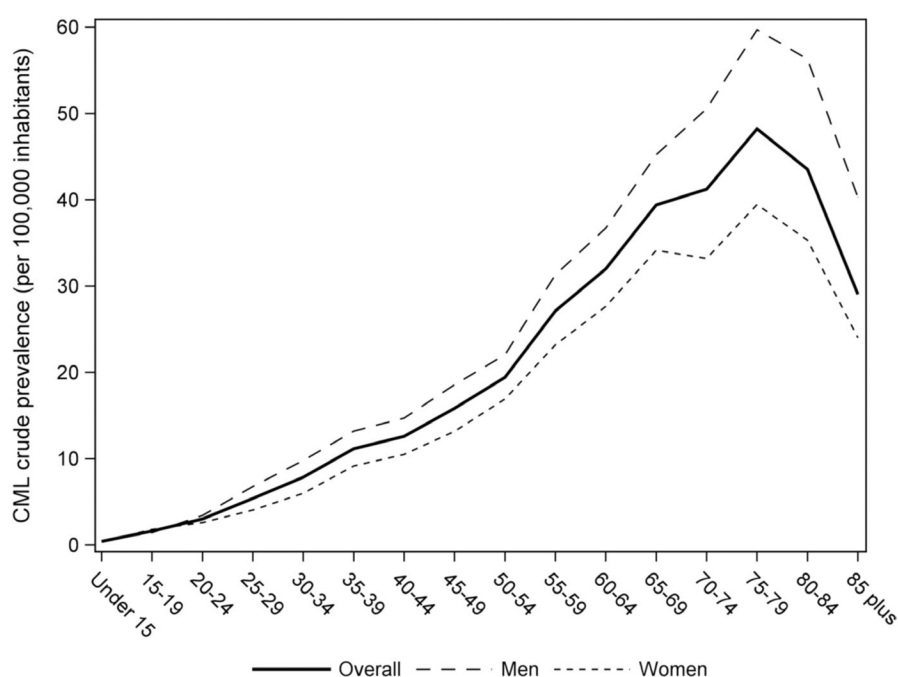
#### 3.2 | Estimation of CML prevalence

On 31 December 2014, the crude prevalence of CML was estimated at 16.3 per 100 000 inhabitants in France [95% confidence interval (CI) 16.0-16.6]. The crude prevalence of CML was 18.5 per 100 000 in men [95% CI 18.0-19.0] and 14.2 per 100,000 in women [95% CI 13.8-14.6]. To allow for international comparisons, standardized prevalence rates are shown in Table 1. CML prevalence increased with age (Figure 2 and Table S2). The crude prevalence of CML was less than 1.6 per 100 000 [95% CI 1.2-2.0] before 20, progressively increased to 19.4 [95% CI 18.1-20.8] at 50-54 years and reached a peak of 48.2 [95% CI 45.4-51.2] at 75-79 years. There was a preponderance of CML in men in all age groups except in younger age groups where prevalence was comparable between genders. The crude prevalence of CML varied from 10.2 to 23.8 per 100 000 inhabitants across all French departments (Figure 3). These variations of prevalence

	Overall	Men	Women
Number of prevalent CML patients on 31 December 2014	10 789	5931	4858
French population in 2014	66 226 643	32 076 965	34 149 678
Crude prevalence of CML per 100 000 inhabitants	16.3 [16.0-16.6]	18.5 [18.0-19.0]	14.2 [13.8-14.6]
Standardized prevalence rates per 100 000 (1960-World population)	10.0 [9.8-10.3]	11.7 [11.4-12.0]	8.4 [8.1-8.6]
Standardized prevalence rates per 100 000 (1976-European population)	13.3 [13.1-13.6]	15.6 [15.2-16.0]	11.1 [10.8-11.4]

Abbreviation: CML: Chronic Myeloid Leukemia.

**TABLE 1** Crude and standardized prevalence rates of CML in France, in 2014



**FIGURE 2** Age-specific and gender-specific crude CML prevalence rates in France in 2014. This figure represents the CML crude prevalence rate by age and sex of the population prevalent on 31 December 2014

persisted to a lesser extent after standardization on gender and age (Table S1).

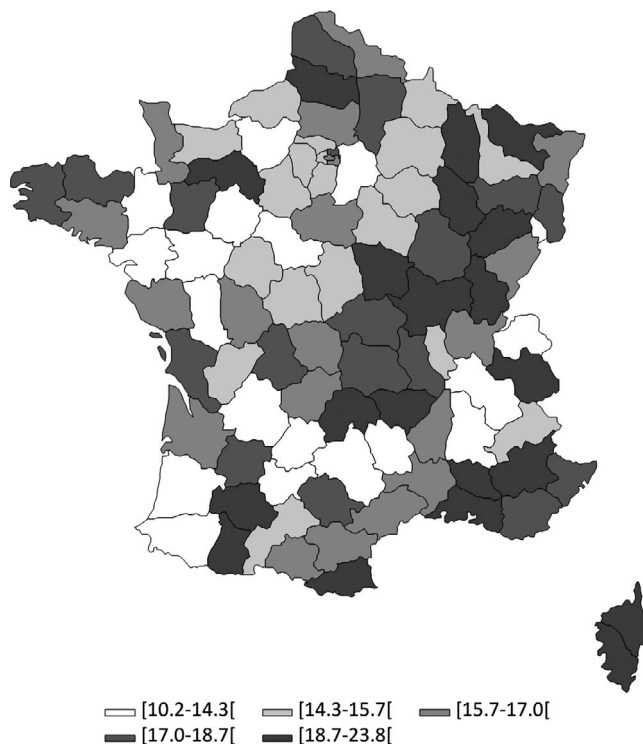
### 3.3 | Algorithm validation to identify CML patients

The comparison of the algorithm with the opinion of hematologists indicated a 96% concordance rate ( $N = 100$ ). According to the hematologists, the algorithm correctly identified 45 true CML patients and correctly excluded 51 false CML patients. However, the algorithm missed three patients who had less than 2 months of TKI reimbursement but were considered as CML patients by the hematologists, based on their sequence of care. Finally, the algorithm selected one false positive CML patient. This was a patient with reimbursement of imatinib for a short time period and missing

data for both hospitalization and eligibility status to full reimbursement of health care expenses for myeloid leukemia.

For the year 2014, the 18 population-based cancer registries from the Francim network identified 150 incident CML patients whereas the algorithm identified 162 new CML patients for the same areas. The median age [interquartile range] of the incident population of CML patients identified using the algorithm was 60 [46-71], with a slight male preponderance (57%). Considering the cancer registries as the gold standard, the algorithm overestimated the number of incident CML patients by 8% in 2014 compared to cancer registries. In 2013 and 2012, the algorithm overestimated the number of incident CML patients by 13% and 14% respectively. In a sensitivity analysis, we provide estimates adjusted using the results of the external validation. The adjusted crude CML prevalence rate was estimated at 15.1 per





**FIGURE 3** CML prevalence rates in 2014 by Department in France. CML: Chronic Myeloid Leukemia. These are crude estimates of CML prevalence per 100 000 inhabitants. Departments are those in which prevalent CML patients lived in 2014

100 000 [95% CI 14.8-15.4] in 2014; 17.1 per 100 000 in men [95% CI 16.7-17.6] and 13.2 per 100 000 in women [95% CI 12.8-13.5].

## 4 | DISCUSSION

Based on the French national health insurance databases, we estimated that the crude prevalence of CML in France was 16.3 per 100 000 inhabitants in 2014. The male preponderance, reported in incidence studies,<sup>2,8,25,26</sup> persisted when studying prevalence and was constant across the ages, except in younger age groups where prevalence was comparable between genders.

Using different sources of data and methodology, our results are consistent with the estimates from Delord et al which ranged from 13.7 per 100 000 inhabitants in 2012 to 17.5 in 2018.<sup>12</sup> CML prevalence estimated by Delord et al was obtained from projections of the French population, incidence rates from six French population-based cancer registries studies and hypotheses on the relative survival of CML patients, while we used individual data from the French national health insurance databases. Our results in 2014 are also consistent with the rising trend in the CML prevalence estimated by Corm et al in the French region Nord-Pas de Calais, from 5.8 per 100 000 inhabitants in

1997 to 10.4 in 2007.<sup>11</sup> Our CML prevalence estimate of 16.3 per 100 000 inhabitants in France in 2014 is between the Swedish estimate (11 per 100 000 inhabitants in 2012 and 14 per 100 000 in 2018)<sup>4,9</sup> and the US prevalence estimates (22.6 per 100 000 in 2010 and 33.5 per 100 000 in 2020).<sup>10</sup> These differences may reflect discrepancies in age distribution, leading to different incidence of CML but also differences in access to TKI with an impact on survival, and differences in methods. It is worth noting that the difference in CML prevalence between Sweden and United States is mainly related to quite different incidence rates (0.90 per 100 000 in Sweden and 1.75 per 100 000 in the United States), which may be due to a wider definition of CML used in the US SEER register or to real differences between the two countries. The increase in CML prevalence may lead physicians to modify their practices in order to care for an increasing number of patients and consider new modalities of follow-up such as clinical routine assessment by clinical nurses. Physicians will also have to adapt to the aging of this population, with a peak of prevalence at 75-79 years in our study, as elderly patients may be more sensitive to side effects of second generation TKIs.

There were variations in CML prevalence across the French territory, which cannot be only explained by differences in age and gender distribution of the population in the French departments (cf. crude and standardized CML prevalence by department Table S1). Data presented here correspond to the distribution of the prevalent CML patients in the departments in which they lived in 2014. These variations are unlikely to be explained by a different quality of data reporting across the departments because information on reimbursement of TKI, on which the algorithm mainly relies, is collected the same way across the French territory. Data regarding environmental exposures are not available in the National Health Insurance databases, precluding association studies using individual data for exposure and outcomes. However, our finding opens research prospects for geographical and ecologic studies.

The French national health insurance databases used in the present study have several advantages: (a) data are collected prospectively and are readily available to be analyzed, (b) the databases cover 98.8% of French population, hence avoiding selection bias and providing a large sample size, (c) data are available from 2006 onwards with no losses to follow-up. A key issue in studies using healthcare administrative databases is to develop and validate an algorithm that identifies patients with the highest accuracy. Different initiatives exist worldwide to encourage the building, validation and use of these algorithms, such as the review of algorithms used to detect various outcomes of interest within the Mini-Sentinel program in the US,<sup>27</sup> and the REDSIAM network in France.<sup>28</sup> Building such algorithms requires expertise in both the studied disease and the administrative databases used.

The analysis of the sequence of care of 200 randomly selected individuals by two expert hematologists using a standardized form was a major step in the building of the algorithm. Indeed, hematologists found a certain amount of miscoding of hematological diagnoses for hospital stays. For example, for a given patient, some hospital discharges were coded CML while it was obvious that the patient suffered from Chronic MyeloMonocytic Leukemia (CMML), another hematological malignancy, given the sequence of care of the patients (treatments, other hospital discharge coded CMML, prognosis...). For that reason, we based our algorithm on the reimbursement of TKIs rather than on hospitalizations with a CML diagnosis code.

To our knowledge, the rare previous studies on CML using healthcare administrative databases<sup>29-31</sup> chose TKI reimbursement combined with claims associated with diagnosis for CML (ICD-9 code: 205.1) to analyze patterns of TKI treatment and healthcare resources consumption without an attempt to quantify prevalence.

Our work has several strengths. Firstly, our algorithm was validated internally by two expert hematologists and externally by comparison with the number of incident CML patients recorded in 18 French population-based cancer registries in the same geographical areas. Secondly, our estimate of CML prevalence is based on individual data at the nationwide level. It also provides estimates of CML prevalence in different geographical areas across France using an identical method of measurement. It can allow comparisons and improve health policy decision-making. Finally, our algorithm could be used to reestimate the CML prevalence rate in France periodically. The algorithm could also be used or adapted in other countries to identify CML patients, provided those countries have databases containing reimbursement/dispensation of drugs data linked with medical discharge diagnosis databases.

Our study also has limitations. The first limitation was the frequent miscoding of hematological diagnoses for hospital stays, which explains why we based our algorithm on the reimbursement of TKIs rather than on hospitalizations with a CML diagnosis code. Hence, the algorithm is unable to identify three categories of true CML patients: (a) patients who have always been treated by interferon alone, (b) patients who have always received TKI in a clinical trial setting, the TKI being provided by the sponsor and thus not reimbursed by the National Health Insurance and (c) elderly patients who have been treated by cytoreductive chemotherapy only. The proportion of patients treated by interferon alone was estimated at 0.7% in a cross-sectional study on the management of first line CML patients in 2013 in France.<sup>32</sup> Regarding the patients who received TKIs in the setting of clinical trials, searching the EU Clinical Trials Register, we found 25 clinical trials starting between 2004 and 2014 in France in CML patients. Excluding trials in which the patients had to be previously

treated by TKI to be included in the trial, we estimated an upper range of 628 patients that could have been missed by the algorithm. Among those patients, several discontinued the trials and were thus identifiable by the algorithm if treated by a TKI outside the trial. In our internal validation, we identified three patients out of 100 who had not received TKI but were considered by the hematologists as possible CML patients given their whole sequences of care. These patients could correspond to the three different mentioned situations. Therefore, the impact of these limitations on the estimation of CML prevalence should be limited.

Secondly, we were unable to validate CML patients identified with our algorithm using medical records because data from the French national health insurance databases for research are anonymous. This difficulty was partially overcome through internal and external validations. However, for the same reason of anonymity, we were unable to perform an individual matching with the registries but we were able to compare total numbers of incident cases between the two sources of data. The algorithm slightly overestimated the number of incident CML patients (from 14% in 2012 to 8% in 2014). Non-CML patients receiving TKI for others diseases than the differential diagnoses already excluded by the algorithm may lead to this slight overestimation. As a sensitivity analysis, we provided adjusted CML prevalence estimates taking into account this possible overestimation. Lastly, the algorithm depends on the care pathway of CML in general and more specifically on the indications of prescription of TKI at the time when we built the algorithm. In the future, some patients in deep molecular remissions may opt for a controlled discontinuation of TKI therapy. They will no longer be treated (and thus identifiable) by TKI. However, before entering into a treatment discontinuation strategy, patients must be treated by TKI for at least 5 years<sup>33</sup> and are therefore identifiable by their TKI intake during this period. To continue to use the algorithm in the coming years, the observation period should be extended as much as possible and start before 2007, year of beginning of inclusion in the first French clinical trial (STIM1) assessing controlled discontinuation of TKI therapy. If new TKIs were to be commercialized, the algorithm should be adapted to take into account this change, by adding new drugs in the list of TKIs.

In conclusion, we built and validated an algorithm to identify CML patients in healthcare administrative databases. We estimated the crude CML prevalence in France at 16.3 per 100 000 inhabitants in 2014, using an innovative methodology based on data from the French national health insurance databases. The algorithm could enable periodical re-estimation of the CML prevalence rate in France and could be implemented in other countries using healthcare claims databases. The algorithm will also offer the opportunity to study



healthcare pathways and healthcare uses of French CML patients or to study the association between adverse events and CML treatments in the French national health insurance databases.

## ACKNOWLEDGMENTS

The authors would like to thank Annie Fourier-Réglat and Catherine Lejeune for scientific expertise, Valérie Edel from the National Institute of Health Data for technical and regulatory advice regarding the data extraction request, Medhi Gabbas and Julien Brand from the French National Health Insurance for data extraction and Zoé Urhy who carried out the data extraction from the cancer registries.

## CONFLICT OF INTEREST

The authors indicate no potential conflicts of interest.

## DATA AVAILABILITY STATEMENT

The data underlying the findings cannot be made freely available because of legal restrictions. Data used for the present study come from the French National Health Insurance databases and include an important number of variables, that, when combined the ones to the others, can lead to re-identifying subjects, and then collecting health information on these persons. This is why the French Data Protection Authority (CNIL) forbids us to make such data freely available. Access to the raw data of the French National Health Insurance must be requested from the National Institute of Health Data (<https://www.indesante.fr/>).

## ORCID

Stéphanie Foulon  <https://orcid.org/0000-0002-0213-0736>

Marc Delord  <https://orcid.org/0000-0002-0455-6749>

## REFERENCES

1. Le Guyader-Peyrou S, Belot A, Maynadié M, et al. Cancer incidence in France over the 1980–2012 period: hematological malignancies. *Rev Epidemiol Sante Publique*. 2016;64(2):103–112. <https://doi.org/10.1016/j.respe.2015.12.017>.
2. Höglund M, Sandin F, Simonsson B. Epidemiology of chronic myeloid leukaemia: an update. *Ann Hematol*. 2015;94(Suppl 2):S241–247. <https://doi.org/10.1007/s00277-015-2314-2>.
3. Apperley JF. Chronic myeloid leukaemia. *Lancet Lond Engl*. 2015;385(9976):1447–1459. [https://doi.org/10.1016/S0140-6736\(13\)62120-0](https://doi.org/10.1016/S0140-6736(13)62120-0).
4. Gunnarsson N, Sandin F, Höglund M, et al. Population-based assessment of chronic myeloid leukemia in Sweden: striking increase in survival and prevalence. *Eur J Haematol*. 2016;97(4):387–392. <https://doi.org/10.1111/ejh.12743>.
5. De Angelis R, Sant M, Coleman MP, et al. Cancer survival in Europe 1999–2007 by country and age: results of EURO CARE–5—a population-based study. *Lancet Oncol*. 2014;15(1):23–34. [https://doi.org/10.1016/S1470-2045\(13\)70546-1](https://doi.org/10.1016/S1470-2045(13)70546-1).
6. Sant M, Minicozzi P, Mounier M, et al. Survival for haematological malignancies in Europe between 1997 and 2008 by region and age: results of EURO CARE–5, a population-based study. *Lancet Oncol*. 2014;15(9):931–942. [https://doi.org/10.1016/S1470-2045\(14\)70282-7](https://doi.org/10.1016/S1470-2045(14)70282-7).
7. Monnereau A, Uhry Z, Bossard N, et al. Survie Des Personnes Atteintes de Cancer En France Métropolitaine, 1989–2013. Partie 2 - Hémopathies Malignes. *Saint-Maurice: Institut de veille sanitaire*. 2016;144.
8. Visser O, Trama A, Maynadié M, et al. Incidence, survival and prevalence of myeloid malignancies in Europe. *Eur J Cancer Oxf Engl*. 2012;48(17):3257–3266. <https://doi.org/10.1016/j.ejca.2012.05.024>.
9. Ohm L, Lundqvist A, Dickman P, et al. Real-world cost-effectiveness in chronic myeloid leukemia: the price of success during four decades of development from non-targeted treatment to imatinib. *Leuk Lymphoma*. 2014;1–7. <https://doi.org/10.3109/10428194.2014.953141>.
10. Huang X, Cortes J, Kantarjian H. Estimations of the increasing prevalence and plateau prevalence of chronic myeloid leukemia in the era of tyrosine kinase inhibitor therapy. *Cancer*. 2012;118(12):3123–3127. <https://doi.org/10.1002/cncr.26679>.
11. Corm S, Micol J, Leroyer A, et al. Kinetic of chronic myeloid leukaemia (CML) prevalence in Northern France since the introduction of imatinib. *J Clin Oncol*. 2008;26(15\_suppl):7088–7088.
12. Delord M, Foulon S, Cayuela J-M, Rousselot P, Bonastre J. The rising prevalence of chronic myeloid leukemia in France. *Leuk Res*. 2018;69:94–99. <https://doi.org/10.1016/j.leukres.2018.04.008>.
13. Gavrilov-Yusim N, Friger M. Use of administrative medical databases in population-based research. *J Epidemiol Community Health*. 2014;68(3):283–287. <https://doi.org/10.1136/jech-2013-202744>.
14. Wiréhn A-BE, Karlsson HM, Carstensen JM. Estimating disease prevalence using a population-based administrative healthcare database. *Scand J Public Health*. 2007;35(4):424–431. <https://doi.org/10.1080/14034940701195230>.
15. Paller AS, Singh R, Cloutier M, et al. Prevalence of psoriasis in children and adolescents in the United States: a claims-based analysis. *J Drugs Dermatol*. 2018;17(2):187–194.
16. Raghu G, Chen S-Y, Yeh W-S, et al. Idiopathic pulmonary fibrosis in US Medicare beneficiaries aged 65 years and older: incidence, prevalence, and survival, 2001–11. *Lancet Respir Med*. 2014;2(7):566–572. [https://doi.org/10.1016/S2213-2600\(14\)70101-8](https://doi.org/10.1016/S2213-2600(14)70101-8).
17. Foulon S, Maura G, Dalichampt M, et al. Prevalence and mortality of patients with multiple sclerosis in France in 2012: a study based on French health insurance data. *J Neurol*. 2017;264(6):1185–1192. <https://doi.org/10.1007/s00415-017-8513-0>.
18. Blin P, Dureau-Pournin C, Foubert-Samier A, et al. Parkinson's disease incidence and prevalence assessment in France using the national healthcare insurance database. *Eur J Neurol*. 2015;22(3):464–471. <https://doi.org/10.1111/ene.12592>.
19. Moulis G, Lapeyre-Mestre M, Palmaro A, Pugnet G, Montastruc J-L, Sailler L. French health insurance databases: what interest for medical research? *Rev Médecine Interne Fondée Par Société Natl Française Médecine Interne*. 2015;36(6):411–417. <https://doi.org/10.1016/j.revmed.2014.11.009>.

20. Tuppinn P, Rudant J, Constantinou P, et al. Value of a national administrative database to guide public decisions: from the système national d'information interrégimes de l'Assurance Maladie (SNIIRAM) to the système national des données de santé (SNDS) in France. *Rev Epidemiol Sante Publique*. 2017;65(Suppl 4):S149-S167. <https://doi.org/10.1016/j.respe.2017.05.004>.
21. Bezin J, Duong M, Lassalle R, et al. The national healthcare system claims databases in France, SNIIRAM and EGB: powerful tools for pharmacoepidemiology. *Pharmacoepidemiol Drug Saf*. 2017;26(8):954-962. <https://doi.org/10.1002/pds.4233>.
22. Estimation de la population au 1<sup>er</sup> janvier | Insee. 2015. <https://www.insee.fr/fr/statistiques/1893198>. Accessed August 17, 2018.
23. Clopper C, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*. 1934;26:404-413. <https://doi.org/10.1093/biomet/26.4.404>.
24. ISD Services | Geography, Population and Deprivation Analytical Support Team | Population | Standard Populations | ISD Scotland. <http://isdscotland.org/Products-and-Services/GPD-Support/Population/Standard-Populations>. Accessed September 4, 2018.
25. Penot A, Preux P-M, Le Guyader S, et al. Incidence of chronic myeloid leukemia and patient survival: results of five French population-based cancer registries 1980-2009. *Leuk Lymphoma*. 2015;1-7. <https://doi.org/10.3109/10428194.2014.974046>
26. Monnereau A, Remontet L, Maynadié M, et al. Estimation nationale de l'incidence des cancers en France entre 1980 et 2012. Partie 2 - Hémopathies malignes. *Inst Veille Sanit*. 2013;88.
27. 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):82-89. <https://doi.org/10.1002/pds.2321>.
28. Goldberg M, Carton M, Doussin A, et al. The REDSIAM network. *Rev Epidemiol Sante Publique*. 2017;65(Suppl 4):S144-S148. <https://doi.org/10.1016/j.respe.2017.06.001>.
29. Chang C-S, Yang Y-H, Hsu C-N, Lin M-T. Trends in the treatment changes and medication persistence of chronic myeloid leukemia in Taiwan from 1997 to 2007: a longitudinal population database analysis. *BMC Health Serv Res*. 2012;12:359. <https://doi.org/10.1186/1472-6963-12-359>.
30. Henk HJ, Woloj M, Shapiro M, Whiteley J. Real-world analysis of tyrosine kinase inhibitor treatment patterns among patients with chronic myeloid leukemia in the United States. *Clin Ther*. 2015;37(1):124-133. <https://doi.org/10.1016/j.clinthera.2014.10.019>.
31. Stenehjem DD, Albright F, Kuo K-L, et al. Response monitoring, tolerability, and effectiveness of imatinib treatment for chronic myeloid leukemia in a retrospective research database. *J Natl Compr Cancer Netw*. 2014;12(8):1113-1121.
32. Etienne G, Huguet F, Guerci-Bresler A, et al. Impact of ELN recommendations in the management of first-line treated chronic myeloid leukaemia patients: a French cross-sectional study. *Br J Haematol*. 2016;174(1):71-80. <https://doi.org/10.1111/bjh.14022>.
33. Mahon F-X. Treatment-free remission in CML: who, how, and why? *Hematol Am Soc Hematol Educ Program*. 2017;2017(1):102-109. <https://doi.org/10.1182/asheducation-2017.1.102>.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Foulon S, Cony-Makhoul P, Guerci-Bresler A, et al. Using healthcare claims data to analyze the prevalence of BCR-ABL-positive chronic myeloid leukemia in France: A nationwide population-based study. *Cancer Med*. 2019;8:3296–3304. <https://doi.org/10.1002/cam4.2200>