Néphrologie & Thérapeutique xxx (xxxx) xxx–xxx

Available online at
ScienceDirect
www.sciencedirect.com

Elsevier Masson France
EM|consulte
www.em-consulte.com

Original article

# An algorithm for identifying chronic kidney disease in the French national health insurance claims database

Imène Mansouri [a,b], Maxime Raffray [c], Mathilde Lassalle [d], Florent de Vathaire [b,e], Brice Fresneau [e,f], Chiraz Fayech [e,f], Hélène Lazareth [g], Nadia Haddy [b,e], Sahar Bayat [c], Cécile Couchoud [c,h,*]

group REDSIAM[1]

[a] *EPI-PHARE (French National Agency for Medicines and Health Products Safety [ANSM] and French National Health Insurance [CNAM]), Saint-Denis, France*
[b] *Center for research epidemiology and population health, Radiation epidemiology team, Université Paris-Saclay, Université Paris-Sud, UVSQ, 94805 Villejuif, France*
[c] *University Rennes, EHESP, REPERES (Recherche en pharmaco-épidémiologie et recours aux soins)–EA 7449, 35000 Rennes, France*
[d] *REIN registry, Agence de la biomédecine, 1, avenue du Stade de France, 93212 Saint-Denis-La Plaine, France*
[e] *Department of children and adolescent oncology, Gustave-Roussy, Université Paris-Saclay, 94805 Villejuif, France*
[f] *Cancer and Radiation, CESP, Unit 1018 Inserm, Villejuif, France*
[g] *Service Évaluation et Outils pour la Qualité et la Sécurité des Soins, Direction de l'Amélioration de la Qualité et de la Sécurité des Soins, Haute Autorité de santé, Saint-Denis, France*
[h] *Université Lyon I, CNRS, UMR 5558, Laboratoire de biométrie et biologie évolutive, équipe biostatistique santé, Villeurbanne, France*

ARTICLE INFO

*Article history:*
Received 4 March 2022
Accepted 11 March 2022

*Keywords:*
Algorithms
Chronic kidney disease
Healthcare claims databases
Validation studies

ABSTRACT

*Background.* – Published algorithms for identifying chronic kidney disease in healthcare claims databases have poor performance except in patients with renal replacement therapy. We propose and describe an algorithm to identify all stage chronic kidney disease in a French healthcare claims databases and assessed its performance by using data from the Renal Epidemiology and Information Network registry and the French Childhood Cancer Survivor Study cohort.
*Methods.* – A group of experts met several times to define a list of items and combinations of items that could be related to chronic kidney disease. For the French Childhood Cancer Survivor Study cohort, information on confirmed chronic kidney disease cases extracted from medical records was considered the gold standard (KDIGO definition). Sensitivity, specificity, and positive and negative predictive value and kappa coefficients were estimated. The contribution of each component of the algorithm was assessed for 1 and 2 years before the start of renal replacement therapy for confirmed end-stage kidney disease in the Renal Epidemiology and Information Network registry.
*Results.* – The algorithm's sensitivity was 78%, specificity 97.4%, negative predictive value 98.4% and positive predictive value 68.7% in French Childhood Cancer Survivor Study cohort and the kappa coefficient was 0.79 for agreement with the gold standard. The algorithm 93.6% and 55.1% of confirmed incident end-stage kidney disease cases from the Renal Epidemiology and Information Network registry when considering 1 year and 2 years, respectively, before renal replacement therapy start.
*Conclusions.* – The algorithm showed good performance among younger patients and those with end-stage kidney disease in the twol last years prior to renal replacement therapy. Future research will address the ability of the algorithm to detect early chronic kidney disease stages and to classify the severity of chronic kidney disease.


* Corresponding author.
*E-mail address:* cecile.couchoud@biomedecine.fr (C. Couchoud).
[1] Other collaborators of the groupe REDSIAM (www.redsiam.fr/): Marie Erbault, Karen Assmann, Nelly Le Guen, Nathalie Poutignat (Haute Autorité de santé), Marine Desbouvries, Sabrina Matz, Lise Thiriet (DRSM Hauts de France), Cyrielle Parmentier (APHP), Marie Metzger, Aghiles Hamroun (Inserm), Philippe Tuppin (CNAM), Fistsum Guebre-Egziabher (HCL), Olivier Moranne (CHU de Nîmes), Isabella Vanorio Vega (ABM).

https://doi.org/10.1016/j.nephro.2022.03.003
1769-7255/

## 1. Abbreviation

| | |
|---|---|
| ATC | Anatomical Therapeutic Chemical |
| CCAM | French Common Classification of Medical Acts |
| CI | confidence interval |
| CKD | chronic kidney disease |
| CKD-EPI | Chronic Kidney Disease-Epidemiology Collaboration |
| CNIL | French Data Protection Authority |
| DCIR | National Health Insurance Claims Database |
| ESKD | end-stage kidney disease |
| FCCSS | French Childhood Cancer Survivor Study |
| eGFR | estimated Glomerular filtration rate |
| ICD-10 | International Statistical Classification of Diseases and Related Health Problems, 10$^{th}$ revision |
| Inserm | National Institute of Medical Research and Health |
| K-coefficient | Cohen kappa coefficient |
| KDIGO | Kidney Disease: Improving Global Outcomes |
| LTFU | long-term follow-up |
| NABM | French Nomenclature of Biological Acts |
| NPV | negative predictive value |
| PMSI | Hospital Discharge Summaries Database |
| PPV | positive predictive value |
| REIN | Renal Epidemiology and Information Network |
| RRT | renal replacement therapy |
| Se | sensitivity |
| SNDS | French administrative healthcare database |
| Sp | specificity |
| TN | true negative |
| TP | true positive |

## 2. Introduction

Chronic kidney disease (CKD) represents a heavy global health burden associated with increased mortality and morbidity and high economic impact [1,2]. The number of individuals with CKD reached more than 700 million in 2017 worldwide, surpassing the number with diabetes mellitus [3,4]. The prevalence of CKD in France is unknown, with some estimates varying between 3 and 6 million, corresponding to about 10% of the French adult population, about 92,000 patients presenting end-stage kidney disease (ESKD) [3,5,6]. Solid data on CKD prevalence in the general population and tools for identifying CKD cases before RRT are lacking and yet, health system planning and policy-making requires careful assessment of CKD epidemiology to develop efficient and cost-effective care strategies.

Health claims databases have long been used to efficiently estimate the prevalence of diseases. These data represent a useful source of information for policy-makers regarding the management of chronic diseases including diabetes, cancer and cardiovascular diseases [7–11]. The analysis of these databases could provide insight into the global burden of CKD and allow for evaluating treatment strategies aimed at slowing its progression. Nevertheless, even though the identification of patients having renal replacement therapy (RRT) in health claims databases is fairly straightforward, identifying other stages of CKD remains challenging.

A systematic review that analyzed several algorithms for CKD based on both diagnostic and procedural codes in 25 administrative databases across 8 countries found poor algorithm performance, yielding low sensitivity and positive predictive value [12]. Another study that identified CKD with diagnostic and procedural codes in Dutch hospital-based database, found higher sensitivity among younger patients and those with advanced CKD [13]. Only a few other studies in Italy and Canada focused on developing algorithms with higher sensitivity based on prescription of specific drugs, medical procedures and hospitalizations related to CKD from healthcare claims data [8,14,15].

In France, the national REIN registry (Renal Epidemiology and Information Network) includes all patients receiving RRT for ESKD. France also has a nationwide health claims database. Unfortunately, results of biological tests (including blood creatinine levels) are not available in this database.

This study aimed to propose and describe an algorithm for the identification of all stage CKD using the French health claims database and assess its performance and utility using data from two different populations: confirmed ESKD cases (REIN registry) one and two years prior RRT and survivors of childhood cancer from the (French Childhood Cancer Survivor Study [FCCSS] cohort).

## 3. Methods

### 3.1. The French administrative healthcare database

The French administrative healthcare database (SNDS) consists of two main databases: the hospital discharge summaries database (PMSI) and the national health insurance claims database (DCIR) and covers 98.8% of the French population, over 66 million persons, from birth (or immigration) to death [16–18]. The PMSI database includes primary, related and associated diagnoses for all private or public medical, obstetric and surgical hospitalizations. These diagnoses are coded according to the International Statistical Classification of Diseases and Related Health Problems, 10th revision (ICD-10) [17]. The date and duration of hospitalization are included. Medical procedures performed during the hospitalization are coded according to the French Common Classification of Medical Acts (CCAM), diagnostic-related groups, as well as highly expensive drugs. The DCIR database includes data on all reimbursed ambulatory care including consultations, medical procedures coded according to the French CCAM, prescribed medications coded according to the Anatomical Therapeutic Classification, and laboratory biological tests coded according to the French Nomenclature of Biological Acts. In addition to including records of all reimbursed ambulatory care, the DCIR contains a list of long-term diseases that allow full reimbursement of costs related to these conditions, with start and end dates. Clinical and biological test results are not available in the database.

The identification of CKD cases in the French SNDS was based on querying all hospital discharge claims, ambulatory care claims, and medication-dispensing data, in private or in public structures.

### 3.2. CKD case definition algorithm

A group of experts in nephrology, renal epidemiology and healthcare claims databases met several times to define a list of items and combinations of items that could be related to CKD. Inclusions of items was made by unanimous decisions. The aggregation of all these items defined the so-called "algorithm". The information on whether a patient may have CKD (identification item) was searched in different components of the SNDS:

- long-term diseases;
- physician claims (consultations);

- drug delivery;
- biological tests;
- diagnosis-related groups;
- hospitalization diagnoses;
- medical acts.

The details of each CKD identification item for each component of the algorithm (nomenclature and codes) are in Supplementary Table 1 for certain (the item is self-sufficient), probable (high probability of being related to CKD, a combination of probable items is required to pass to the certain level) and possible (a combination of possible items is required to pass to the probable level) CKD cases.

Certain items consisted of:

- hospitalization with at least one diagnosis of CKD: ICD-10 codes N00-N08 (glomerular diseases), N11 and N13-N16 (renal tubulo-interstitial diseases), N18 (CKD), E102 (type 1 diabetes with diabetic CKD), E112 (type 2 diabetes with diabetic CKD), T861 (kidney transplant failure and rejection), Z49 (care involving dialysis) and Z940 (kidney transplantation);
- at least two consultations with a nephrologist during one calendar year;
- combinations of prescribed medications used in treating CKD including erythropoiesis-stimulating agents, drugs for treating hyperkalemia and hyperphosphatemia, angiotensin-converting enzyme inhibitors, iron, antacids with sodium bicarbonate, vitamin D, calcium, high doses of diuretics and hepatitis B vaccine with the specialty of the prescriber;
- medical acts involving RRT by dialysis or kidney transplantation and creation of arteriovenous fistula;
- different combinations of biological tests involved in the diagnosis and/or follow-up of CKD: creatinine clearance, complete blood electrolytes, blood urea nitrogen, parathyroid hormone blood test, serum S-25-hydroxyvitamin D, hepatitis B surface antibody dosage and urine testing for protein.

The association of at least two of the following probable items also led to a certain identification of CKD:

- other hospitalization probably related to CKD with the following diagnostic codes: I13 (hypertensive heart and renal disease), I151 (hypertension secondary to other renal disorders), N171 (acute renal failure with medullary necrosis), N280 (ischemia and infarction of kidney) and Q61 (cystic kidney disease);
- other medication delivery related to CKD prescribed by a nephrologist;
- biological tests related to CKD prescribed by a nephrologist;
- medication delivery for CKD prescribed by a nephrologist and with different medical procedures related to the creation or the surgical repair of arteriovenous fistula, renal biopsy and arterial Doppler imaging.

### 3.3. Study population and data sources

#### 3.3.1. French Childhood Cancer Survivor Study cohort

The French Childhood Cancer Survivor Study cohort (FCCSS) cohort includes 7670 5-year childhood cancer survivors who received treatment from 1942 to 2000 for solid cancers or lymphomas before age 21 in several French centers. Among them, 4567 treated at Gustave Roussy Hospital alive in 2012 were eligible for a long-term follow-up (LTFU) visit. A total of 1002 (22%) attended the long-term follow-up (LTFU) between 2012 and 2018. Systematic screening offered by the LTFU clinic included clinical examination and urine and biological testing [19]. Serum levels of creatinine and markers of kidney damage (proteinuria, hematuria, calcium, phosphate, glycosuria, phosphorus reabsorption rate, etc.) were reported in medical records. CKD was defined according to the Kidney Disease: Improving Global Outcomes definition as functional abnormalities (tubulopathies, proteinuria...) of the kidney regardless of estimated glomerular filtration rate (eGFR) or eGFR $< 60$ mL/min/1.73 m$^2$ [20]. GFR was estimated according to the Chronic Kidney Disease Epidemiology Collaboration equation [21]. Partial nephrectomy without functional consequence was not considered as CKD. All cases were confirmed by an expert. Information on the diagnosis of any CKD was extracted from medical records and considered the gold standard of confirmed CKD. A total of 867 childhood cancer survivors from the FCCSS cohort with at least one LTFU visit had available outpatient data in the SNDS (Supplementary Fig. 1).

The FCCSS protocol was approved by the Inserm national ethics committee and the French National Agency regulating Data Protection (Cnil no. 902 287). Consent was obtained from patients, parents or guardians according to national research ethics requirements.

#### 3.3.2. The REIN registry: confirmed ESRD adult patients with RRT

Since 2012, the REIN registry has gathered data on all new ESKD patients who started RRT in metropolitan France and its overseas territories. The registry includes data on patient identification (age, sex, and postcode of the place of residence), comorbidities (e.g., cardiovascular diseases, diabetes, cancer), and characteristics at RRT start (eGFR, hemoglobin and serum albumin levels, planned or emergency dialysis, center identification, etc.) [22]. Patients are followed, and specific events, such as placement on a waitlist for kidney transplantation, kidney transplantation and death, are recorded. To obtain information on patients' healthcare consumption before RRT, this population was linked to the SNDS by using a deterministic and iterative linkage method that was previously described [23]. Two years of healthcare consumption data prior RRT were extracted for all adults ($\geq$18 years old) with ESKD who were included in the REIN registry and started RRT in France in 2015. Data for long-term diseases were not available for these patients.

### 3.4. Statistical analysis

After specifying the algorithm, it was locked and following analysis were performed.

Medical data from the subset of the FCCSS cohort who attended at least one LTFU visit at Gustave Roussy LTFU Clinic and with data from the French SNDS were compared at an individual level. The algorithm was applied, including different combinations of codes, with medical records as the gold standard for determining a CKD case. The final identification of patients with CKD in the SNDS was based on items considered certain and/or the combinations of at least two probable items.

To evaluate the performance of the algorithm in identifying diagnosed CKD, different indicators were calculated: sensitivity (Se), specificity (Sp), positive predictive value (PPV), negative predictive value (NPV), accuracy and Cohen's kappa coefficient (k-coefficient) and their 95% confidence interval (CI). Se was calculated as the proportion of cases classified as positive by both the algorithm and medical record review, or "true positives" (TPs), as compared with all CKD cases identified by the gold standard (medical record review). Sp was calculated as the proportion of cases without CKD identified by both the algorithm and the gold standard, or "true negative" (TNs), as compared with all negative cases by the gold standard. PPV was calculated as the proportion of

TPs divided by all potential CKD cases identified by the algorithm and medical record review. NPV was defined as the number of TNs divided by the number of patients with a negative classification for CKD by the algorithm and the gold standard.

For the population of confirmed ESKD cases extracted from the REIN registry, the sensibility of the algorithm was calculated as the proportion of cases identified by the algorithm compared to the total number of ESRD cases recorded by the REIN registry (gold standard). The final classification of patients with CKD in the SNDS was based on items considered certain, probable or possible.

The algorithm was used with the available SNDS healthcare data separately for 1 and 2 years before RRT start. We then assessed the contribution of each component of the algorithm by using Venn diagrams [24].

All statistical analyses involved using SAS 9.4.

## 4. Results

### 4.1. Validation of the algorithm in the subset of FCCSS cohort with LTFU visit

In the FCCSS cohort, 1002 patients had an LTFU visit and available data on renal function at the date of the visit; 135 were excluded because of pairing failure with the health insurance database (Supplementary Fig. 1). The characteristics of the validation population ($n = 867$) were compared to those of the excluded population. The groups did not significantly differ in type of primary cancer malignancy. However, the validation sample was slightly younger (median age 35.4 [IQR 2.8–49.7]) and more frequently had a diagnosis of primary childhood malignancy in recent years or hypertension than the excluded population (Supplementary Table 2). When the validation cohort ($n = 867$) was compared to the 3535 excluded patients who never had a LTFU visit, females, CNS (Central Nervous System) tumor survivors and patients with comorbidities showed up more for LTFU visits.

A total of 59 childhood cancer survivors had CKD confirmed by clinicians during the LTFU visit including 4 ESKD, detailed description is shown in Supplementary Table 3. In the French administrative healthcare database (SNDS), among them, 29 (49.2%) cases were coded as certain with the algorithm due to a hospitalization diagnoses, 25 (42.4%) were coded as certain because of physician claims and 21 (35.6%) by long-term illness exemption due to severe or chronic nephropathy (Supplementary Table 4).

A total of 67 patients were considered as CKD by the algorithm (at least one certain items or at least 2 probable items). Therefore, for identifying confirmed CKD cases (all stages), in the FCCS cohort, the algorithm Se was 78% (95% CI 67.4–88.5), Sp 97.4% (95% CI 96.3–98.5), NPV 94.8% (87.5–99.3) and PPV 68.7% (57.6–79.8) (Table 1). Concerning level of agreement with the gold standard, the *k*-coefficient for the algorithm was 0.79 (95% CI 0.61–0.80). When restricting the analysis to survivors of nephroblastoma (Wilm's tumors) ($n = 127$), both sensitivity and specificity remained similar, at 78.4% (95% CI 65.1–91.6) and 97.6% (95% CI 96.4–98.7) respectively. Analysis of false-negative and false-positive cases are provided in Supplementary Tables 5 and 6. The sensitivity was significantly higher with our algorithm compared to the used of hospital claims alone due to reclassification of false negative (Table 1).

### 4.2. Sensitivity of the algorithm with the confirmed ESKD cases from the REIN registry

Among the 11,083 patients from the REIN registry who started RRT in 2015 in France, data for 9627 (86.8%) were linked with the

**Table 1**
Test characteristics of the CKD case definition algorithms applied in the French SNDS using confirmed CKD cases in the subset of FCCSS cohort with at least one LTFU visit as the gold standard.

| | N | TP | FP | TN | FN | Se (%) (95% CI) | Sp (%) (95% CI) | PPV (%) (95% CI) | NPV (%) (95% CI) | Acc (%) (95% CI) | k (%) (95% CI) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Based only on hospital claims | 867 | 29 | 19 | 789 | 30 | 49.2 (36.4–61.9) | 97.6 (96.6–98.7) | 60.4 (46.6–74.3) | 96.3 (95.1–97.6) | 94.3 (92.8–95.9) | 0.51 (0.39–0.63) |
| Identification algorithm | 867 | 46 | 21 | 787 | 13 | 78.0 (67.4–88.5) | 97.4 (96.3–98.5) | 68.7 (57.6–79.8) | 98.4 (97.5–99.3) | 96.1 (94.8–97.4) | 0.79 (0.61–0.80) |
| Identification algorithm excluding survivors of renal malignancies tumor | 740 | 29 | 17 | 686 | 8 | 78.4 (65.1–91.6) | 97.6 (96.4–98.7) | 63.0 (49.1–77.0) | 98.3 (97.3–99.2) | 96.6 (95.3–97.9) | 0.75 (0.59–0.90) |
| Identification algorithm among survivors of renal malignancies | 127 | 17 | 4 | 101 | 5 | 77.3 (59.8–94.8) | 96.2 (92.5–99.8) | 81.0 (64.7–97.8) | 95.3 (91.3–99.3) | 92.9 (88.4–97.4) | 0.75 (0.59–0.90) |

FCCSS: French Childhood Cancer Survivor Study; LTFU: long-term follow-up; *n*: total number of subjects included in the validation sample; TP: true positive; FP: false positive; Se: sensitivity; Sp: specificity; PPV: positive predictive value; NPV: negative predictive value; Acc: accuracy; *k*: Cohen's kappa coefficient; 95% CI: confidence interval.

**Table 2**
Proportion of confirmed ESKD incident cases in 2015 in France identified by the algorithm in the SNDS according to the time frame considered before renal replacement therapy (RRT) start.

| Component of the algorithm and case status | Adult patients starting RRT in 2015 in France (confirmed ESKD cases), $n = 9493$ | |
|---|---|---|
| | 1 year before RRT, $n$ (%) | 2 years before RRT, $n$ (%) |
| Physician claims (visit) | | |
| Certain | 6410 (67.5) | 4088 (43.1) |
| Probable | 1262 (13.3) | 1323 (13.9) |
| Possible | 533 (5.6) | 1161 (12.2) |
| Undetected | 1288 (13.6) | 2921 (30.8) |
| Medication deliverance | | |
| Certain | 4925 (51.9) | 2268 (23.9) |
| Probable | 2167 (22.8) | 2124 (22.4) |
| Possible | 1563 (16.5) | 3579 (37.7) |
| Undetected | 838 (8.8) | 1522 (16) |
| Biological tests | | |
| Certain | 2374 (25) | 1613 (17) |
| Probable | 3351 (35.3) | 2252 (23.7) |
| Possible | 3111 (32.8) | 4113 (43.3) |
| Undetected | 657 (6.9) | 1515 (16) |
| Diagnoses-related groups | | |
| Certain | 4576 (48.2) | 642 (6.8) |
| Probable | 1232 (13) | 333 (3.5) |
| Undetected | 3685 (38.8) | 8518 (89.7) |
| Medical acts | | |
| Certain | 6307 (66.4) | 701 (7.4) |
| Probable | 1119 (11.8) | 270 (2.8) |
| Possible | 481 (5.1) | 988 (10.4) |
| Undetected | 1586 (16.7) | 7534 (79.4) |
| Hospitalization diagnoses | | |
| Certain | 6778 (71.4) | 1494 (15.7) |
| Probable | 50 (0.5) | 32 (0.3) |
| Possible | 22 (0.2) | 24 (0.3) |
| Undetected | 2643 (27.8) | 7943 (83.7) |
| Total | | |
| Certain | 8885 (93.6) | 5226 (55.1) |
| Probable | 240 (2.5) | 959 (10.1) |
| Possible | 261 (2.7) | 2381 (25.1) |
| Undetected | 107 (1.1) | 927 (9.8) |

CKD: chronic kidney disease; ESRD: end-stage renal disease.

SNDS; 134 did not have any healthcare consumption in the SNDS database before RRT start and were considered inherently undetectable by the algorithm and thus were excluded from this validation analysis. Hence, 9493 patients were included in the analysis (Supplementary Fig. 2).

The algorithm identified 8885 (93.6%) of the confirmed incident ESKD cases from the REIN registry as certain cases when considering 1 year before RRT start. Only 107 (1.1%) confirmed cases were not identified as cases.

The period considered was of importance: the algorithm identified 5526 (55.1%) confirmed cases as certain cases when considering the 2 years before RRT start (Table 2).

The proportion of certain cases identified by the hospitalization diagnoses and diagnosis-related groups components of the algorithm greatly increased between the 2 periods. Fig. 1 shows the evolution of cases identified by the algorithm according to their status (certain, probable, possible) between 2 years and 1 year before RRT start. Most probable, possible and previously undetected cases were identified as certain cases in the year before RRT start (95%, 87% and 81%, respectively). The contribution of each component is presented in Fig. 2.

## 5. Discussion

Accurately identifying patients with CKD at the national level is an ambitious challenge in CKD epidemiology but ultimately critical in healthcare policy-making and evaluation. Using healthcare databases is a promising perspective. The algorithm based on healthcare data we present in this article is a first and important step toward this goal. With validation in 2 different populations and contexts, we show good performance of this algorithm.

Previous algorithms have been developed to detect CKD patients in healthcare claims databases. Some algorithms benefit from serum creatinine results, which are of great value to detect CKD cases; an example is the Alberta Kidney Disease Network (AKDN) database, which combined administrative databases with laboratory data [8]. However, serum creatinine value is lacking in many healthcare claims databases, including the French health insurance databases. Other authors used algorithms based on diagnosis at hospital discharge to identify CKD: this was the case for the 16 studies included in a systematic review published in 2010 [12]. In such selected populations, Sp is high but Se is poor. This approach is not conclusive in evaluating the burden of CKD in the general population because it selects only hospitalized patients, who may not be representative. The performance of 11 diagnostic codes and their combination was analyzed with 7 databases in Ontario, Canada [15]. The results showed high Sp but low Se, especially in early-stage CKD. In our study, Se was lower when only hospital claims were used. Only one recent algorithm used information on medications and outpatient services combined with that from a hospital discharge registry and a ticket exemption registry [14]. This algorithm identified 99.457 individuals with CKD (mean age 71 years, 55.8% males). The exclusive contributions of each regional source were 35.047 (35.2%) from the

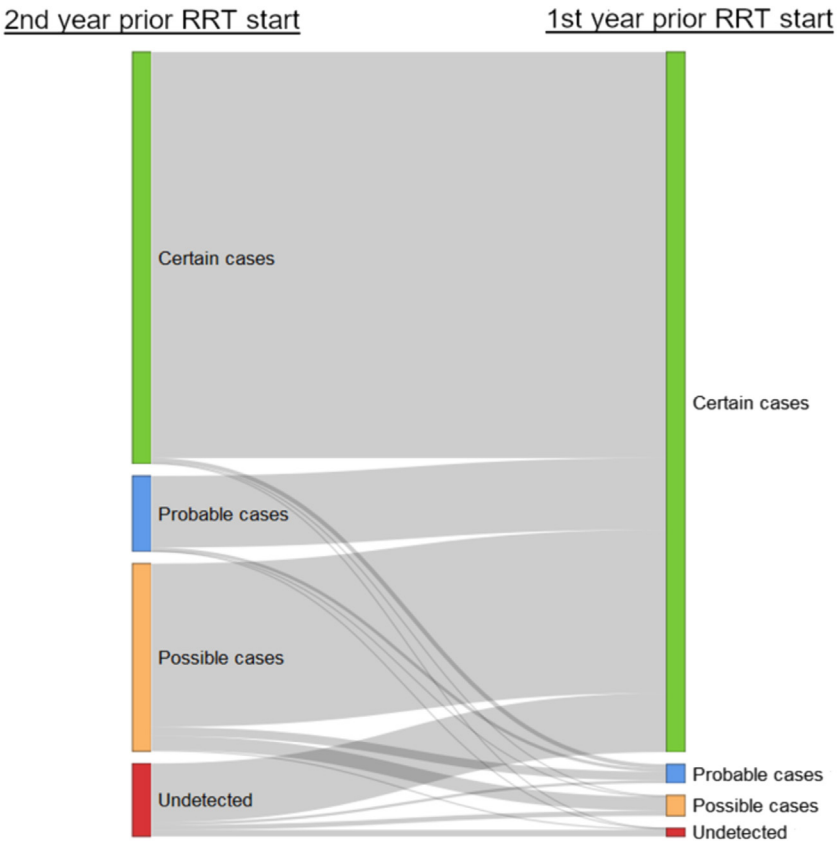


**Fig. 1.** Flow of confirmed ESRD new cases identified by the algorithm between the second and first year before RRT start. The contribution of the algorithm to the identification of certain cases differed across its components (Fig. 2). When considering the year before RRT, 406 (4.5%) cases were identified solely by the Medical acts component. Conversely, the biological tests and diagnosis-related groups identified only 26 and 9 cases, respectively. Two years before RRT, the consult and medication components identified 1317 (14.8%) and 403 (4.5%) of certain cases, respectively. ESRD: end-stage renal disease; RRT: renal replacement therapy.

Outpatient specialist Service information system, 27.778 (27.9%) from the hospital discharge registry, 4143 (4.2%) from the ticket exemption registry and 463 (0.5%) from a drug dispensing registry; 5.1% of cases were found in all databases. However, because of the lack of a gold standard, this algorithm was validated in only dialysis patients.

The low performance of these algorithms to correctly identify CKD patients (TP rate) may be due to a high number of false negatives. Indeed, CKD remains a silent disease for a long time and associated with non-specific symptoms, so its diagnosis is difficult for health professionals. This situation could explain the lack of specific healthcare consumption until advanced stages of CKD and for some patients close to RRT as well by some quality issues in coding. Our algorithm showed good performance in the FCCSS cohort, with Se $>$ 70% and Sp $>$ 97%. False negatives were mostly patients with CKD stage 2 and renal tumor (Supplementary Table 2). The high sensitivity found in the FCCSS could be explained by the inclusion of younger patients (median age 35.4 years old). Similar results were also reported in a study based on the Dutch hospital-based database [13]. Also, CKD in this population could be related to risk factors different from those in the general population. Nevertheless, the diagnosis and management of CKD were based on the KDIGO guidelines as in the general population Second, the LTFU guidelines for survivors undergoing unilateral renal surgery recommends an annual assessment of renal function, which may lead to a possible over-diagnosis bias of CKD in survivors of renal malignancies [25]. Furthermore, only 22% of survivors included in the FCCSS cohort alive in 2012 had an LTFU visit and a renal function assessment; 75.4% (49 patients) of those with confirmed CKD received the diagnosis during the LTFU visit. This observation emphasizes the crucial role of this visit in the LTFU of childhood cancer survivors.

Because the REIN registry ESRD cohort consisted of only ESRD patients (i.e., no negative cases), Se could not be estimated. However, this analysis allows for showing that our algorithm is performant in more severe disease stages, close to RRT. Indeed, the algorithm correctly identified 93.6% confirmed incident cases of ESRD during the year before RRT. Of note, the 107 (1.1%) confirmed cases not identified as cases by the algorithm represent patients with late referral and without any health consumption before RRT. As RRT drew near, medical acts and hospital diagnosis were more prominent as sources of identification. During the 2 years before RRT start, medications and visits with a nephrologist were more frequent sources of identification.

Our study may suffer from the following limits. The validation was made in a selected small cohort that may not be representative of the general population. The classification of the items in certain, probable or possible is rather subjective and may be discussed. Sensitivity analyses are planned.

Finally, our algorithm was developed by a group of experts and not data-driven. Although it seems to present good performance as is, this methodological challenge of CKD identification is an iterative process and will be updated regularly. For example, the pool of items classified as indicative of possible cases of CKD (Supplementary Table 1) was not used here to identify patients with CKD and represent an area of further research.

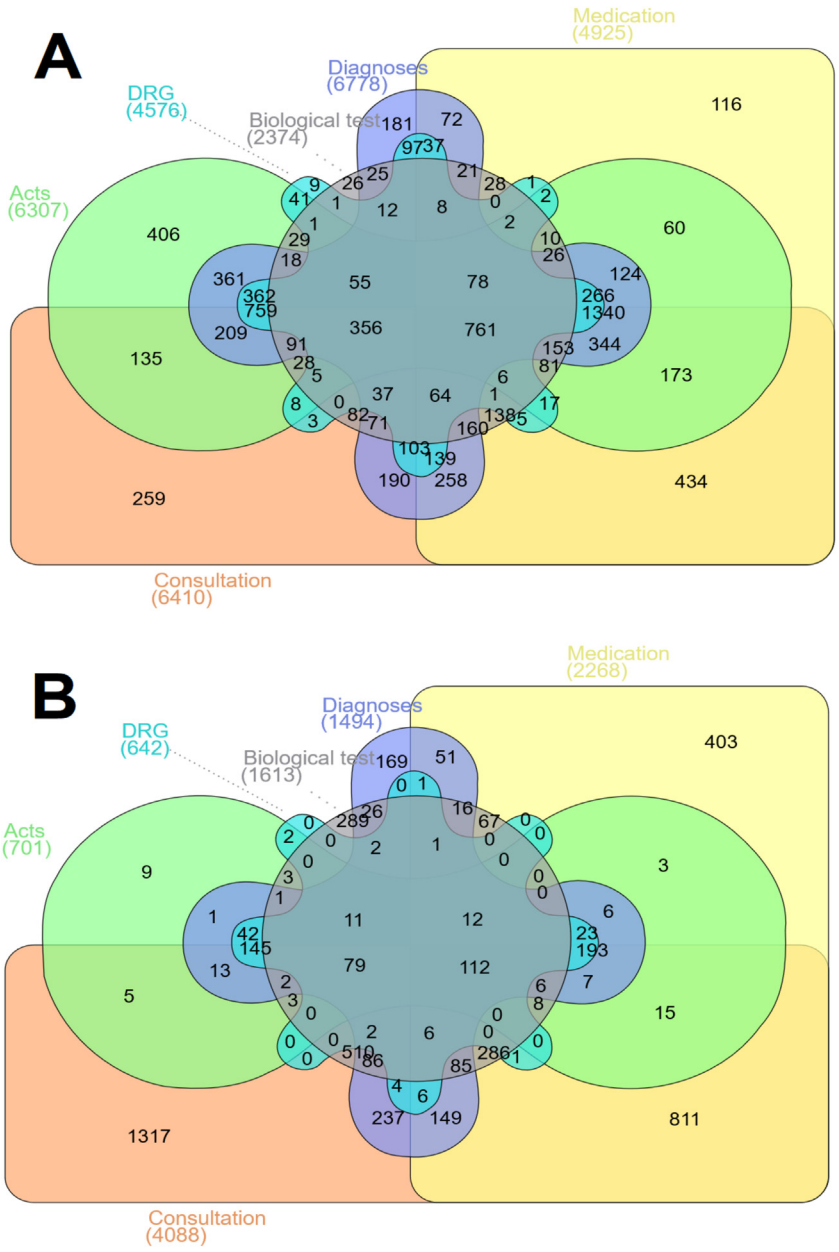


**Fig. 2.** Origin of the cases identified as certain in the confirmed end-stage kidney disease (ESKD) population according to the component of the algorithm (consultation, medication delivery, biological tests, medical acts, hospitalization diagnoses, diagnosis-related groups [DRG]) in the first (A) and second (B) year before renal replacement therapy start ($n$ = 8885).

Nevertheless, identifying milder stages of CKD can be challenging because it requires more complex and advanced case-finding algorithms. Future research will address the ability of the algorithm to detect all CKD stages and classify individuals at early, advanced or late stage of CKD as well as the use of other populations and contexts for further validation.

## 6. Conclusion

Our algorithm showed good performance among young patients and those with ESKD in the two last years prior to RRT. Because it is not based on lab results, it can be used in various contexts, especially in big medico-administrative databases. Further improvements and other validations in various populations are planned.

## Ethics approval and consent to participate

The present study is based on secondary use of previous studies. All participants of these two studies were informed of the possibility of secondary used of their data.

## Authors' contributions

IM, MR, ML, HL, NH, SB and CC worked on the expert-driven algorithm.

IM, FV, BF, CF and NH validated the algorithm in the FCCSS cohort.

MR and SB validated the algorithm in the REIN cohort.

IM, MR and CC wrote the manuscript.

All the authors reviewed the manuscript and made valuable contribution.

**Funding**

The FCSS study has received some funds from The Gustave Roussy Foundation.

**Disclosure of interest**

The authors declare that they have no competing interest.

## Acknowledgements

The authors wish to thank all REIN registry and FCCSS participants.

## Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.nephro.2022.03.003.

## References

[1] Foreman KJ, Marquez N, Dolgert A, Fukutaki K, Fullman N, McGaughey M, et al. Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality for 250 causes of death: reference and alternative scenarios for 2016-40 for 195 countries and territories. Lancet 2018;392:2052–90.
[2] Nichols GA, Ustyugova A, Déruaz-Luyet A, O'Keeffe-Rosetti M, Brodovicz KG. Health care costs by type of expenditure across eGFR stages among patients with and without diabetes, cardiovascular disease, and heart failure. J Am Soc Nephrol 2020;31:1594–601.
[3] Bikbov B, Purcell CA, Levey AS, Smith M, Abdoli A, Abebe M, et al. Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2020;395:709–33.
[4] James SL, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018;392:1789–858.
[5] Haute Autorité de santé (HAS). Maladie renale chronique. Guide du parcours de soins; 2012.
[6] Couchoud C, Lassalle M. Rapport REIN Registry—2019 Annual Report. Agence de la Biomédecine; 2021, http://www.agence-biomedecine.fr/IMG/pdf/rapport_rein_2019_2021-06-23.pdf.
[7] Lipscombe LL, Hwee J, Webster L, Shah BR, Booth GL, Tu K. Identifying diabetes cases from administrative data: a population-based validation study. BMC Health Serv Res 2018;18:316.
[8] Bello A, Hemmelgarn B, Manns B, Tonelli M, Alberta Kidney Disease Network. Use of administrative databases for health-care planning in CKD. Nephrol Dial Transplant 2012;27(Suppl. 3):iii12–8.
[9] Saczynski JS, Andrade SE, Harrold LR, Tjia J, Cutrona SL, Dodd KS, et al. A systematic review of validated methods for identifying heart failure using administrative data: detection of congestive heart failure in claims. Pharmacoepidemiol Drug Saf 2012;21:129–40.
[10] Choma NN, Griffin MR, Huang RL, Mitchel EF, Kaltenbach LA, Gideon P, et al. An algorithm to identify incident myocardial infarction using Medicaid data. Pharmacoepidemiol Drug Saf 2009;18:1064–71.
[11] Ajrouche A, Estellat C, De Rycke Y, Tubach F. Evaluation of algorithms to identify incident cancer cases by using French health administrative databases: cancer incidence in the sniiram/egb databases. Pharmacoepidemiol Drug Saf 2017;26:935–44.
[12] Vlasschaert MEO, Bejaimal SAD, Hackam DG, Quinn R, Cuerden MS, Oliver MJ, et al. Validity of administrative database coding for kidney disease: a systematic review. Am J Kidney Dis 2011;57:29–43.
[13] van Oosten MJM, Brohet RM, Logtenberg SJJ, Kramer A, Dikkeschei LD, Hemmelder MH, et al. The validity of Dutch health claims data for identifying patients with chronic kidney disease: a hospital-based study in the Netherlands. Clin Kidney J 2021;14:1586–93.
[14] Marino C, Ferraro PM, Bargagli M, Cascini S, Agabiti N, Gambaro G, et al. Prevalence of chronic kidney disease in the Lazio region, Italy: a classification algorithm based on health information systems. BMC Nephrol 2020;21:23.
[15] Fleet JL, Dixon SN, Shariff SZ, Quinn RR, Nash DM, Harel Z, et al. Detecting chronic kidney disease in population-based administrative databases using an algorithm of hospital encounter and physician claim codes. BMC Nephrol 2013;14:81.
[16] Moulis G, Lapeyre-Mestre M, Palmaro A, Pugnet G, Montastruc JL, Sailler L. French health insurance databases: what interest for medical research? Rev Med Int 2015;36:411–7.
[17] Tuppin P, de RL, Weill A, Ricordeau P, Merliere Y. French national health insurance information system and the permanent beneficiaries sample. Rev Epidemiol Sante Publique 2010;58:286–90.
[18] Bezin J, Duong M, Lassalle R, Droz C, Pariente A, Blin P, et al. The national healthcare system claims databases in France, SNIIRAM and EGB: powerful tools for pharmacoepidemiology. Pharmacoepidemiol Drug Saf 2017;26:954–62.
[19] Haghiri S, Fayech C, Mansouri I, Dufour C, Pasqualini C, Bolle S, et al. Long-term follow-up of high-risk neuroblastoma survivors treated with high-dose chemotherapy and stem cell transplantation rescue. Bone Marrow Transplant 2021, http://www.nature.com/articles/s41409-021-01258-1.
[20] Stevens PE. Evaluation and management of chronic kidney disease: synopsis of the kidney disease: Improving Global Outcomes 2012 Clinical Practice Guideline. Ann Intern Med 2013;158:825.
[21] Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro III AF, Feldman HI, et al. A new equation to estimate glomerular filtration rate. Ann Intern Med 2009;150:604–12.
[22] Couchoud C, Stengel B, Landais P, Aldigier JC, de CF, Dabot C, et al. The renal epidemiology and information network (REIN): a new registry for end-stage renal disease in France. Nephrol Dial Transplant 2006;21:411–8.
[23] Raffray M, Bayat S, Lassalle M, Couchoud C. Linking disease registries and nationwide healthcare administrative databases: the French renal epidemiology and information network (REIN) insight. BMC Nephrol 2020;21:25.
[24] Heberle H, Meirelles GV, da Silva FR, Telles GP, Minghim R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. BMC Bioinformatics 2015;16:169.
[25] Long Term Follow-up Guidelines For Survivors of Childhood, Adolescent And Young Adult Cancers. Version 5.0; 2018, http://www.survivorshipguidelines.org/pdf/2018/COG_LTFU_Guidelines_v5.pdf.