Clinical Kidney Journal, 2024, vol. 17, no. 5, sfae117

https://doi.org/10.1093/ckj/sfae117
Advance Access Publication Date: 23 April 2024
Original Article

ORIGINAL ARTICLE

Prevalence of chronic kidney disease in France: methodological considerations and pitfalls with the use of Health claims databases

Cécile Couchoud [1], Maxime Raffray [2], Mathilde Lassalle [1], Zhanibek Duisenbekov [1], Olivier Moranne [3], Marie Erbault [4], Hélène Lazareth [5], Cyrielle Parmentier [6], Fitsum Guebre-Egziabher [7], Aghiles Hamroun [8], Marie Metzger [9], Imene Mansouri [10], Marcel Goldberg [11], Maris Zins [11], Sahar Bayat-Makoei [2] and Sofiane Kab [11]; on behalf of the group REDSIAM « kidney disease »

[1]Réseau Epidémiologie et Information en Néphrologie, Agence de la Biomédecine, Saint-Denis-La-Plaine, France, [2]Univ. Rennes, EHESP, CNRS, Inserm, Arènes - UMR 6051, RSMS (Recherche sur les Services et Management en Santé) - U 1309 – Rennes, France, [3]Service Néphrologie-Dialyse-Apherese, Hôpital Universitaire Caremau, Nîmes, IDESP Université de Montpellier, France, [4]Haute Autorité de Santé, Saint-Denis-La-Plaine, France, [5]Service de Néphrologie, HEGP, APHP, Paris, France, [6]Service de Néphrologie, Trousseau, APHP, Paris, France, [7]Service Néphrologie-Dialyse-Aphérèse-Hypertension, Hôpital Edouard Herriot, Hospices Civils de Lyon, Université Lyon-1 INSERM U 1060, Lyon, France, [8]Department of Public Health – Epidemiology, Department of Nephrology, Lille University Hospital Center, RIDAGE, Pasteur Institute of Lille, Inserm, Lille University, Lille, France, [9]Center for Research in Epidemiology and Population Health, Paris-Saclay University, Paris-Sud University, Versailles Saint Quentin University, Inserm, Villejuif, France, [10]Direction Procréation, Embryologie et Génétique Humaine, Agence de la Biomédecine, Saint-Denis-La-Plaine, France and [11]Cohorte CONSTANCES, Inserm UMS11, Villejuif, France

Correspondence to: Cécile Couchoud; E-mail: cecile.couchoud@biomedecine.fr

ABSTRACT

Background. Health policy-making require careful assessment of chronic kidney disease (CKD) epidemiology to develop efficient and cost-effective care strategies. The aim of the present study was to use the RENALGO-EXPERT algorithm to estimate the global prevalence of CKD in France.

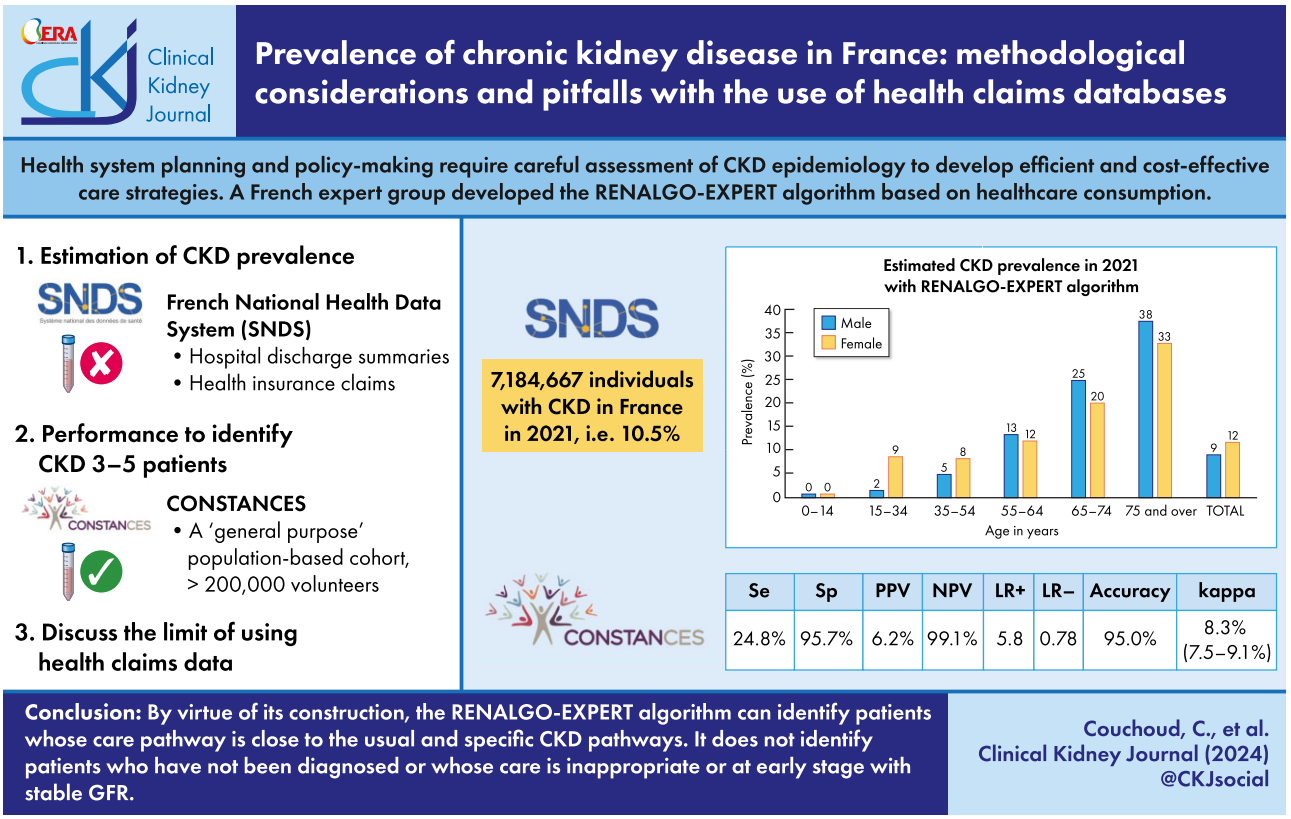
Methods. An expert group developed the RENALGO-EXPERT algorithm based on healthcare consumption. This algorithm has been applied to the French National Health claims database (SNDS), where no biological test findings are available to estimate a national CKD prevalence for the years 2018–2021. The CONSTANCES cohort (+219 000 adults aged 18–69 with one CKD-EPI eGFR) was used to discuss the limit of using health claims data.

Received: 19.12.2023; Editorial decision: 16.3.2024

Results. Between 2018 and 2021, the estimated prevalence in the SNDS increased from 8.1% to 10.5%. The RENALGO-EXPERT algorithm identified 4.5% of the volunteers in the CONSTANCES as CKD. The RENALGO-EXPERT algorithm had a positive predictive value of 6.2% and negative predictive value of 99.1% to detect an eGFR<60 ml/min/1.73 m². Half of 252 false positive cases (ALGO+, eGFR > 90) had been diagnosed with kidney disease during hospitalization, and the other half based on healthcare consumption suggestive of a 'high-risk' profile; 95% of the 1661 false negatives (ALGO−, eGFR < 60) had an eGFR between 45 and 60 ml/min, half had medication and two-thirds had biological exams possibly linked to CKD. Half of them had a hospital stay during the period but none had a diagnosis of kidney disease.

Conclusions. Our result is in accordance with other estimations of CKD prevalence in the general population. Analysis of diverging cases (FP and FN) suggests using health claims data have inherent limitations. Such an algorithm can identify patients whose care pathway is close to the usual and specific CKD pathways. It does not identify patients who have not been diagnosed or whose care is inappropriate or at early stage with stable GFR.

GRAPHICAL ABSTRACT



Keywords: CKD, diabetes, epidemiology, health claims databases, hypertension

KEY LEARNING POINTS

What was known:

- The prevalence of total CKD is unknow, due to a lack of CKD registries in many European countries. Some recent estimates varying between 7% and 10% of the adult population have been reported.
- To overcome the limitations of having to rely on access to repeated determinations of eGFR, health claims databases have been used in various countries to estimate the prevalence of CKD.

This study adds:

- The estimated prevalence of CKD in France at 8%–10% is close to that expected.
- By virtue of its construction, the RENALGO-EXPERT algorithm can identify patients whose care pathway is close to the usual and specific CKD pathways. It does not identify patients who have not been diagnosed or whose care is inappropriate or at early stage with stable GFR.

Potential impact:

- Health system planning and policy-making require careful assessment of CKD epidemiology to develop efficient and cost-effective care strategies that aimed at slowing its progression.
- This algorithm will now be used by French Health authorities as a contributing tool for CKD burden assessment and optimizing care delivery. However, complementary tools will have to be associated with this approach to address its inherent limits.

INTRODUCTION

Chronic kidney disease (CKD) represents a heavy global health burden associated with increased mortality and morbidity and high economic impact. Worldwide, in 2017, 697.5 million (95% UI 649.2 to 752.0) cases of all-stage CKD were recorded, for a global prevalence of 9.1% (8.5 to 9.8) [1]. The prevalence of kidney failure with replacement therapy in Europe and France is well known, thanks to the European ERA registry [2] and the French National Renal Epidemiology and Information Network (REIN) registry [3–5]. However, the prevalence of total CKD is unknown, due to lack of CKD registries in many European countries [6]. Some recent estimates varying between 7% and 10% of the adult population have been reported [7]. A precise assessment of CKD epidemiology is critical for sustainable and efficient planning (e.g. resource allocation) and to develop, implement, and evaluate cost-effective policies aimed at controlling CKD [8].

One main difficulty in identifying CKD patients is the silent nature of the disease in its early stages, with non-specific symptoms in the more advanced stages. Its identification is complicated by the definition of this disease, critically reliant on biological results on a given period [9]. The definition of CKD includes all individuals with markers of kidney damage or those with an estimated glomerular filtration rate (eGFR) of $<$60 ml/min/1.73 m^2 on at least two occasions 90 days apart (with or without markers of kidney damage). Markers of kidney disease may include: albuminuria (albumin: creatinine ratio ACR $>$3 mg/mmol), haematuria (or presumed or confirmed kidney origin), electrolyte abnormalities due to tubular disorders, kidney histological abnormalities, structural abnormalities detected by imaging (e.g. polycystic kidneys, reflux nephropathy) or a history of kidney transplantation. Then, CKD is classified based on the eGFR and the level of albuminuria and allow risk stratification of the patients. Patients are classified as G1–G5 on the basis of the eGFR, and A1–A3 on the basis of the urine ACR [10].

To overcome the limitations of having to rely on access to repeated determinations of eGFR, health claims databases have been used in various countries to estimate the prevalence of CKD [11–18]. As a product of its universal health system coverage, France possesses one of the largest nationwide claims databases in the world, the National Health Data System (SNDS), covering the entire French population, i.e. 67 million inhabitants [19]. A joint effort of clinicians and researchers, referred to as the 'REDSIAM Kidney Disease' group, has led to the development of an algorithm called RENALGO-EXPERT that aims to identify patients with CKD using the claims data available in the SNDS [20].

The primary objective of this study was to estimate the prevalence of CKD across France by applying the RENALGO-EXPERT algorithm to the National Health Claims Database (SNDS). A secondary objective was to assess the performance of this algorithm and explore the challenges associated with using healthcare databases for prevalence estimation. For that secondary objective, we used the CONSTANCES population-based cohort that is linked to the SNDS database and includes eGFR data.

MATERIALS AND METHODS

The CKD case definition algorithm: RENALGO-EXPERT algorithm

The method and results of the first version of our CKD case definition algorithm has been published previously [20]. Briefly, a consortium of experts in nephrology, kidney epidemiology, and healthcare claims databases collaborated to design a practical algorithm for identifying CKD cases. This algorithm evaluates CKD likelihood through a combination of indicators associated with the CKD care pathway. These indicators encompass various components of the French National Health Data System (SNDS), including, but not limited to, chronic health conditions, nephrologist consultations, CKD relevant medications, CKD relevant biological tests, CKD relevant medical acts, hospitalization records with CKD-linked diagnoses, and CKD-related diagnostic groups. Inclusions of items were made by unanimous decisions. In the initial step, each item was categorized into three categories: 'certain', 'likely', or 'possible' CKD item. The subsequent step involves classifying persons in two groups as 'certain' or 'likely' based on their holistic care pathway (recurring and combined health claims compatible with a CKD). This comprehensive approach offered a framework for CKD identification and classification. The validation of this algorithm in the French Childhood

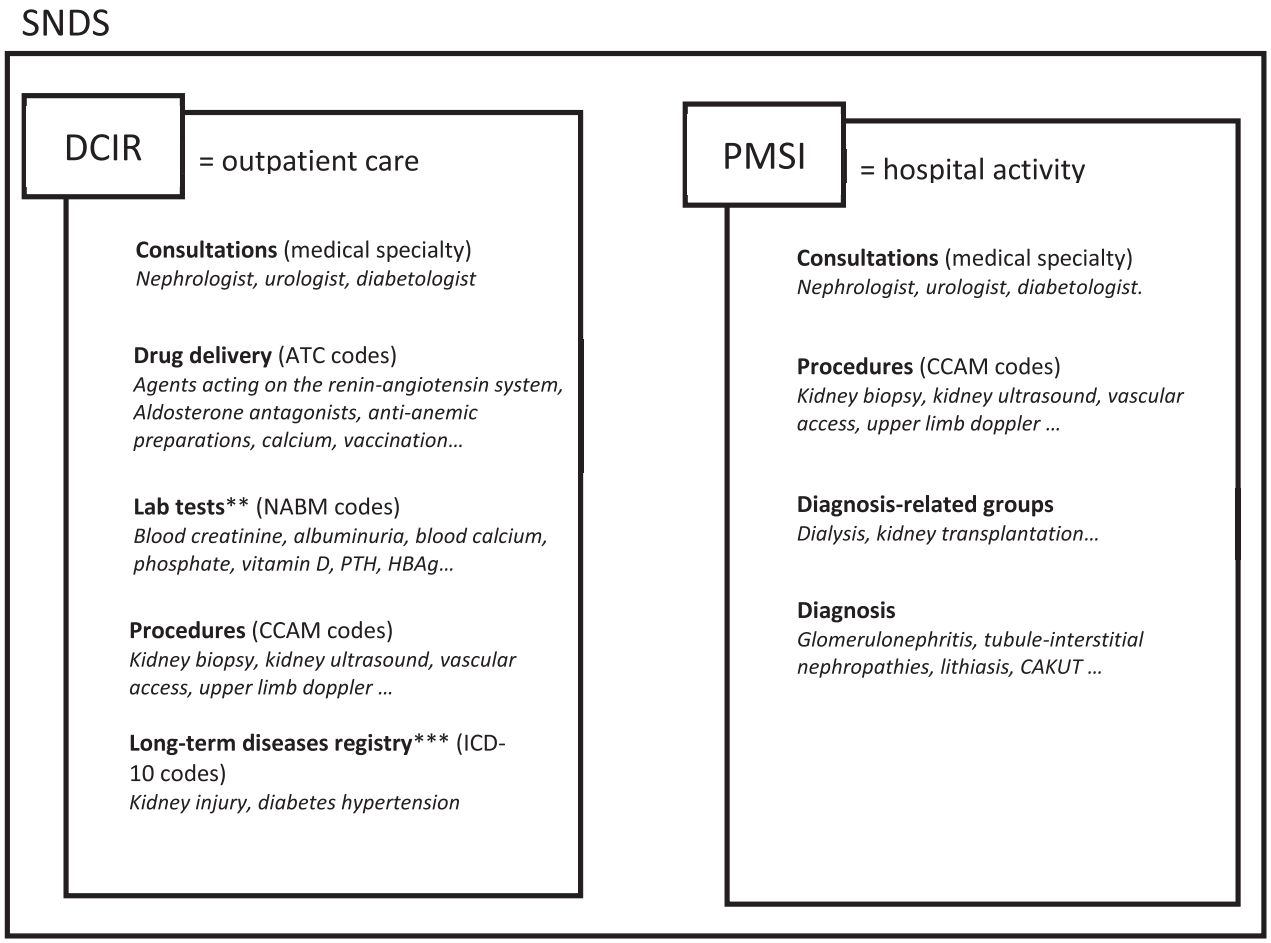


Figure 1: Overview of the French SNDS database and the data of interest for RENALGO-EXPERT algorithm[a].

Cancer Survivor Study cohort showed good performance with a sensitivity >70% and a specificity of >97% [20]. This first version was however improved by adding the notion of repeated claims over a 1-year period and additional combination between items.

Data sources

The French administrative healthcare database (SNDS)

The SNDS consists of two main databases: the Hospital Discharge Summaries Database (PMSI) and the National Health Insurance Claims Database (DCIR). It covers an extensive 98.8% of the French population, which translates to >66 million individuals spanning from birth (or immigration) to death [21–23]. The value of the SNDS rely on its national coverage, its comprehensiveness, the information provided at the individual level, and its regular updating. Figure 1 presents an overview of the SNDS, in particular applied to our subject. Within the PMSI database, a comprehensive array of primary, related, and associated diagnoses are catalogued for either private or public medical, obstetric, and surgical hospitalizations. These diagnoses adhere to the coding system outlined in the *International Statistical Classification of Diseases and Related Health Problems, 10th Revision* (ICD-10) [17]. This repository also features details such as hospitalization dates and durations. Furthermore, it encompasses encoded data about medical procedures conducted during hospital stays, coded according to the French Common Classification of Medical Acts (CCAM), as well as diagnosis-related groups and notably expensive pharmaceuticals. The DCIR database, on the other hand, encapsulates information on all reimbursed ambulatory care, consultations, medically coded procedures based on the French CCAM, prescribed medications categorized by the Anatomical Therapeutic Classification, and laboratory biological tests following the French Nomenclature of Biological Acts (NABM). Alongside its comprehensive coverage of reimbursed ambulatory care, the DCIR provides a compilation of chronic conditions warranting full reimbursement for related costs, supplemented by initiation and termination dates (*Affection longue durée*). However, it is worth noting that clinical and biological test findings are not encompassed within this database. Therefore, no information can be found considering CKD stages. Compared to some healthcare claims databases of other countries (USA or Sweden, for example), it is worth noting that the French

SNDS data for outpatient visits do not contain diagnoses (main or contributory).

THE CONSTANCES COHORT

The CONSTANCES cohort is a 'general purpose' population-based epidemiological cohort that started in 2012. It is a French nationally representative sample of >200 000 volunteers aged between 18 and 69 at inclusion [24]. It aims at contributing to the development of epidemiological research and to provide public health information. In addition to the baseline and annual self-administrated questionnaire completed at home, subjects underwent a health examination used to collect health-related data: clinical examination, blood analysis, blood pressure, weight, height and waist-to-hip ratio, electrocardiogram, spirometry, sight, and hearing examination. Systematic albuminuria and proteinuria detection were only implemented in 2018. Active follow-up is ensured by a postal self-questionnaire to be completed every year at home, and an invitation for a follow-up visit every 4–5 years is planned for all cohort volunteers.

CONSTANCES' volunteers are also followed by annual direct linkage with the SNDS using Social Security Number.

In this cohort, the kidney condition was evaluated by the glomerular filtration rate (eGFR) estimated for each CONSTANCES volunteer using the CKD-EPI formula based on serum creatinine measured enzymatically at inclusion. Therefore, only CKD stage 3–5 was considered. Those enrolled before 2017 had the opportunity to provide a second eGFR measure and had similar characteristics in terms of age, sex, comorbidities, and eGFR values compared to the complete population (Supplementary Table S1).

Statistical analyses

Estimation of the overall French CKD prevalence

The RENALGO-EXPERT algorithm was applied to the SNDS data from the years 2018 to 2021. The prevalence of CKD in the French population was estimated as the number of people classified with the algorithm as CKD 'certain' or 'likely' divided by the total of the individuals with claims in the given year. Among patients coded as 'certain', those with identified claims related to dialysis or kidney transplantation were distinguished as such. No information can be given considering CKD stage since it does not include any information on eGFR or urinary results.

Performance of the algorithm

In the CONSTANCES cohort, the eGFR was estimated for each volunteer using the CKD-EPI formula, based on the blood creatinine measurement available at inclusion. CKD was defined as an eGFR below 60 ml/min/1.73 m². In parallel, based on health claims from the year before inclusion, the volunteers were classified using the RENALGO-EXPERT algorithm into 'without CKD', 'certain CKD', or 'likely CKD'.

To evaluate the performance of the RENALGO-EXPERT algorithm in identifying stage 3–5 CKD volunteers, different indicators were calculated: sensitivity (Se), specificity (Sp), positive predictive value (PPV), negative predictive value (NPV), accuracy and the Cohen's kappa coefficient (*k*-coefficient), and its 95% confidence interval (CI).

The performance of the algorithm was also assessed using two other eGFR thresholds: 30 and 90 ml/min/1.73 m².

Subgroup analyses were done among persons that had a second measure of serum creatinine at the fifth year of follow-up and among those with a test of albuminuria or proteinuria. In these groups, CKD was defined as two values of eGFR <60 ml/min/1.73 m² or albuminuria >30 mg/mmol. Subgroup analyses were also performed in individuals with higher risk of CKD, i.e. with known diabetes or hypertension at inclusion.

Qualitative analysis of 'diverging' cases

Volunteers with eGFR >90 ml/min/1.73 m² but classified as CKD certain with the RENALGO-EXPERT algorithm and volunteers with an eGFR <60 ml/min/1.73 m² but classified as 'without CKD' by the RENALGO-EXPERT algorithm were explored. Their clinical characteristics and health claims were reviewed.

All statistical analyses involved was carried out using SAS v.9.4.

Data availability and ethics

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. All procedures have been approved by the Institutional Review Board of the French Institute of Health (Inserm) (Opinion no. 01–011, then no. 21–842), and authorized by the by the French Data Protection Authority (Commission Nationale de l'Informatique et des Libertés, CNIL) (Authorization #910486). The biobank obtained a favourable opinion from the Committee for the protection of individuals: CPP Sud Est I (#2018–32) and an authorization from the CNIL (#DR-2–2018-137). All volunteers sign a written consent form for their participation in CONSTANCES, and, where applicable, for their participation in the biobank.

RESULTS

Prevalence of CKD in the French population (SNDS database)

In 2018, 5 459 509 individuals were identified by RENALGO-EXPERT algorithm as having CKD, 5 521 404 in 2019, 6 584 667 in 2020, and 7 184 667 in 2021, representing an estimated prevalence in the general population increasing from 8.1% to 10.5% (Table 1). Prevalence increased with age and sex ratio varied among age groups (Fig. 2). The individuals identified as 'certain' CKD represented 13% of the total of individuals identified by RENALGO-EXPERT algorithm. The characteristics of the patients did not vary according to the years. The median age was 67 years (Q1–Q3: 51–78) [66 years (Q1–Q3: 49–77) in the 'likely' group, 74 years (Q1–Q : 61–85), in the 'certain' group, 68 years (Q1–Q3: 55–77), in the 'certain' group with RRT]. There were 54% women (56% in the 'likely' group, 42% in the 'certain' group, 35% in the 'certain' group with RRT). There were 18% of individuals treated for diabetes, 16% with antihypertensive drugs and 7% with lipid-lowering agents. By only including individuals aged 15–74 years, the estimated prevalence in 2021 would have been 9.4%.

Performance of the algorithm in the CONSTANCES cohort

In the CONSTANCES cohort, 196 647 volunteers out of 206 278 were linked to the SNDS database and had at least one serum creatinine measurement available (Supplementary Fig. S1). The

Table 1: Prevalence of CKD in the French population: number of individuals identified by the RENALGO-EXPERT algorithm in the SNDS National Health Database[a].

	2018		2019		2020		2021	
RENALGO-EXPERT classification	Number of individuals identified as CKD	Prevalence (denominator 67 296 443)	Number of individuals identified as CKD	Prevalence (denominator 67 749 000)	Number of individuals identified as CKD	Prevalence (denominator 68 344 615)	Number of individuals identified as CKD	Prevalence (denominator 68 713 200)
CKD Certain with Renal replacement therapy	75 943	0.11%	80 789	0.12%	79 427	0.12%	81 695	0.12%
CKD Certain without Renal replacement therapy	658 993	0.98%	698 722	1.03%	726 473	1.06%	719 229	1.05%
CKD likely	4 724 573	7.02%	4 741 893	7.00%	5 778 767	8.46%	6 383 743	9.29%
Total	5 459 509	8.11%	5 521 404	8.15%	6 584 667	9.63%	7 184 667	10.46%

[a]No information can be given considering CKD stage because SNDS does not include any information on eGFR or urinary results.

non-included individuals were marginally younger and had a slightly higher women representation, with an average age of 45 years (SD 13.8) compared to 47 years (SD 13.5) for those included in the analysis. The percentage of women in the not-included group was 58.6%, versus 53.4% in the included group. Based on the eGFR at inclusion, 34.8% of them had an eGFR <90 ml/min/1.73 m² and 1.1% <60 ml/min/1.73 m².

Based on the RENALGO-EXPERT algorithm, 4.5% of the volunteers in the CONSTANCES had CKD, 790 and 8025 volunteers as CKD 'certain' and CKD 'likely', respectively. The characteristics of the population, according to this classification are described in Table 2. Volunteers identified as CKD with the algorithm are older and have more often associated cardiovascular risk factors (i.e. diabetes, hypertension, or dyslipidaemia).

The RENALGO-EXPERT algorithm had a sensitivity (Se) of 24.8%, specificity (Sp) of 95.7%, PPV of 6.2%, and NPV of 99.1% to detect CKD defined by eGFR <60 ml/min/1.73 m² (Table 3). Accuracy was high (95%), mainly because of the proportion of true negative due to the low prevalence. The kappa coefficient indicated a moderate agreement.

By increasing the eGFR reference value for CKD definition, the proportion of false negative increased from 0.8% (ref. 60 ml/min/1.73 m²) to 32.9% (ref. 90 ml/min/1.73 m²), while the proportion of false positive decreased from 4.2% (ref. 60 ml/min/1.73 m²) to 2.6% (ref. 90 ml/min/1.73 m²) (Table 4). By decreasing the eGFR reference value, the proportion of false negatives decreased from 0.8% (ref. 60 ml/min/1.73 m²) to 0.04% and 0.0% (refs 45 and 30 ml/min/1.73 m², respectively), while the proportion of false positive increased from 4.2% (ref. 60 ml/min/1.73 m²) to 4.4% 4.5% (ref. 45 and 30 ml/min/1.73 m², respectively). The RENALGO-EXPERT algorithm's sensitivity at the 30 ml/min/1.73 m² definition was 91.8%.

Subgroup analysis

During the follow-up, 57 308 volunteers (28%) had a second measure of serum creatinine (mean follow-up = 5.2 years). Among them, 26 593 (46.4%) had an eGFR ≥90 ml/min/1.73 m² at inclusion and at follow-up, 22.3% had a decrease of their eGFR (Supplementary Table S1). In this subgroup, the RENALGO-EXPERT algorithm identified 3.8% of the volunteers with CKD. The RENALGO-EXPERT algorithm had a Se of 20.1%, Sp of 96.5%, PPV of 3.7%, and NPV of 99.1% to detect CKD defined by eGFR <60 ml/min/1.73 m². (Supplementary Table S2).

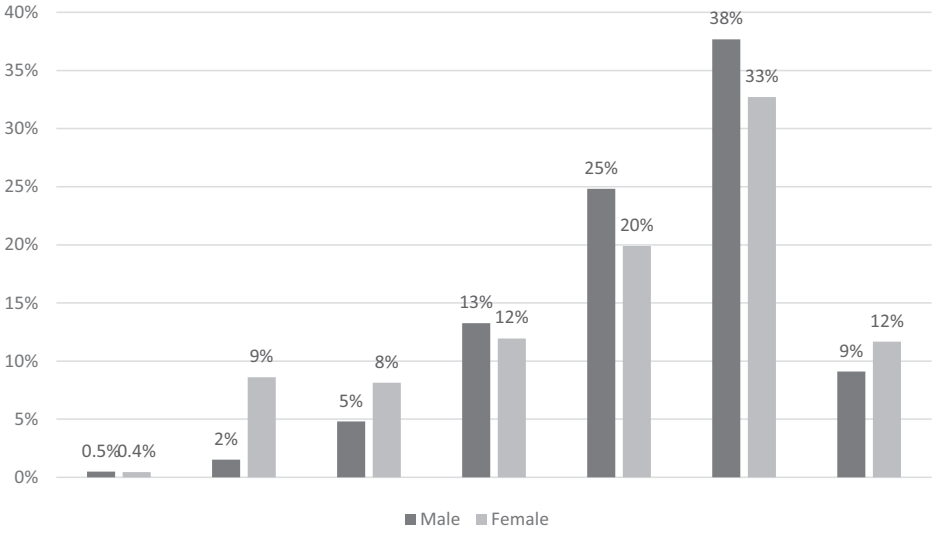
Among the 196 647 volunteers who participated, 21 356 were enrolled between 2019 and 2020. Among this group, 16 593 underwent testing for both albuminuria/proteinuria and creatinine levels in the same examination and 14 579 had two serum creatinine measurements. Only 28 individuals had both two measures of eGFR <60 ml/min/1.73 m², and/or proteinuria >0.5 g/l, or ACR >30 mg/mmol (Supplementary Table S3).

At baseline, 5716 volunteers had diabetes and 26 974 hypertension (Supplementary Table S4). In these subgroups, sensitivity increased as well as the PPV and kappa coefficient with a slight decrease of the specificity and NPV at the cost of increased false positives (Table 5). Accuracy decreased more in volunteers with diabetes (78.5%) than in volunteers with hypertension (87.2%).

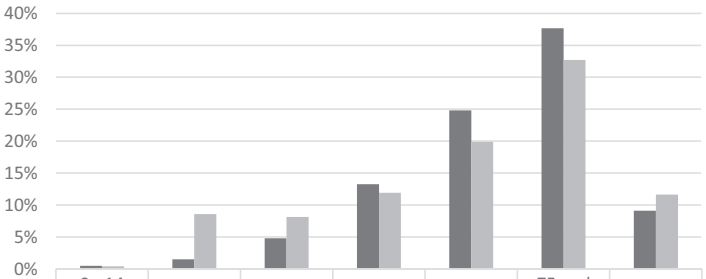
Sensitivity predictive values and PPVs were higher in younger volunteers and in men.

Qualitative analysis of 'diverging' cases

In total, 252 volunteers with eGFR>90 ml/min/1.73 m² were classified as CKD 'certain' with the algorithm (false positive). Half



Estimated CKD prevalence in 2021 with RENALGO-EXPERT algorithm



	0 - 14 years	15 - 34	35- 54	55 - 64	65 - 74	75 and over	TOTAL
% Men	0.5%	1.5%	4.8%	13.2%	24.8%	37.7%	9.1%
% Women	0.4%	8.6%	8.1%	11.9%	19.9%	32.7%	11.7%
Men with CKD	29 466	124 907	413 821	553 374	878 814	1035 649	3036 031
Women with CKD	24 807	713 189	721 402	526 457	794 888	1341 650	4122 393

Figure 2: Prevalence of CKD identified by RENALGO-EXPERT in 2021 according to age and gender in the SNDS y.
[a]No information can be given considering CKD stage because SNDS does not include any information on eGFR or urinary results.

Table 2: Characteristics of the CONSTANCES cohort patients according to their classification with the RENALGO-EXPERT algorithm.

N	CKD CERTAIN 790	CKD LIKELY 8025	NO CKD 187 832	Total 196 647
Age, mean (SD)	54.9 (12.2)	49.7 (14.3)	47.0 (13.4)	47.2 (13.5)
% Male	54.7	36.9	46.9	46.5
% with hypertension	71.1	44.1	29.8	30.5
% with diabetes	19.1	15.3	3.5	4.0
% with dyslipidaemia	58.5	45.1	31.4	32.0
eGFR (ml/min/1.73 m²)				
Q1	55.1	82.1	85.6	85.4
Median	77.7	94.0	95.8	95.7
Q3	94.1	106.5	106.3	106.3
% eGFR < 90	68.1	39.8	34.4	34.8
% eGFR < 60	30.2	3.8	0.9	1.1
% eGFR < 45	15.3	0.8	0.04	0.13
% eGFR < 30	6.2	0.1	0.0	0.03

of them had a kidney disease declared in the Health Claim Database (ICD10 at hospital discharge), which may have been diagnosed in the presence of a proteinuria with a normal kidney function or with fluctuant GFR, which would explain that at inclusion in the CONSTANCES cohort they had an eGFR >90 ml/min/1.73 m². The other half were classified because of a combination between compatible biological exams and drugs, which may also correspond to 'high profiles' such as volunteers with diabetes or cardiovascular comorbidities.

A total of 1661 volunteers were not classified as CKD but had an eGFR <60 (false negative), and 95% of them had an eGFR between 45 and 60 ml/min and a mean age of 63 years at inclusion in the CONSTANCES cohort. Half had medication and two-thirds had biological exams possibly related to CKD. Half of them had a hospital stay during the period but none had a diagnosis of kidney disease.

Table 3: Performance of the RENALGO-EXPERT algorithm according to the level of eGFR at 60 ml/min/1.73 m².

	eGFR CONSTANCES cohort	
RENALGO-EXPERT Algorithm	eGFR < 60	eGFR ≥ 60
CKD +	TP = 548 (0.3%)	FP = 8 267 (4.2%)
CKD −	FN = 1661 (0.8%)	TN = 186 171 (94.7%)

TP, true positive; FN, false negative; FP, false positive; TN, true negative; eGFR in ml/min/1.73 m²;

Se	Sp	PPV	NPV	LR+	LR-	Accuracy	Cohen's kappa coefficient
24.8%	95.7%	6.2%	99.1%	5.8	0.78	95.0%	**8.3% (07.5%–9.1%)**

Se, sensitivity; Sp, specificity,

DISCUSSION

Based on a complex algorithm that includes various combinations of health care claims, the prevalence of CKD in the general population of France is estimated at 8%–10%, using national health claims data SNDS with no biological values. This result is in accordance with other estimations of CKD prevalence in the general population. Using the CONSTANCES cohort that has eGFR data available, the RENALGO-EXPERT algorithm showed moderate performance in identifying CKD stages 3–5. In fact, the analysis of diverging cases suggests that this approach has limitations that should be borne in mind.

CKD prevalence according to available sources

Unlike our approach, many published studies had at their disposal measurement of serum creatinine and/or urine albumin in a sample of the general population allowing them to use the internationally validated definition of CKD. However, all these results have to be interpreted in the light of the characteristics of the sample, the definition of chronicity, the choice of a given equation for estimating GFR and the method of measuring serum creatinine or albuminuria [25–29]. In France, the MONALISA study using three representative cross-sectional surveys in subjects aged 35–74.9 years estimated the CKD prevalence at 8.2% (95% CI, 7.4–8.9%) [30]. In Europe, the adjusted CKD stages 1–5 prevalence in the adult population has been reported between 3.31% (95% CI, 3.30 to 3.33) in Norway and 17.3% (95% CI, 16.5 to 18.1) in the Northeast German Study of Health in Pomeranzia study [31]. In the USA, with the data from the National Health and Nutrition Examination Surveys (NHANES 2015–2018), the prevalence of CKD was estimated at 13.3% (95% CI, 12.3%–14.4%) [32]. In China, the overall prevalence of CKD was estimated at 10.8% (10.2–11.3) [33]. In an adult Arabic-Berber population in Morocco, the adjusted prevalence of CKD was estimated at 5.1% [34]. In Canada, in individuals managed in primary care, 7.4% were identified as having CKD [35]. A systematic review of 100 studies comprising 6 908 440 patients, reported a global prevalence of 13.4% for CKD stages 1–5 and 10.6% for CKD stages 3–5 [36].

Although access to biological results is the preferred method, this is often not possible on a large scale. Therefore, other studies have used diagnoses coded in health claims databases to estimate the prevalence of CKD. They did not directly use the results of biological tests but the diagnosis coded by health professionals. On the basis of diagnosis at hospital discharge, in the Ontario study, 7.7% of the patients were classified as positive for the CKD database algorithm using 11 ICD codes [12]. The US Renal Data System Coordinating Center identifies patients with CKD in administrative data sets by using diagnosis codes from inpatient claims or at least two from outpatient claims or physician and supplier service claims for kidney disease and comorbid conditions [37]. CKD prevalence was estimated at ∼7% of the Medicare population.

In the CaReMe CKD study, using both measured and diagnosed CKD from digital healthcare systems in 11 countries, the pooled prevalence of possible CKD was 10.0% (95% CI 8.7–11.4), defined as having a CKD diagnosis or one pathological UACR or eGFR value, where the chronicity of CKD was not confirmed. When using two pathological UACR or eGFR values at least 90 days apart the estimation was 7.0% (5.6–8.5), with only one value it raised to 9.0% (7.6%–10.4%) When using a registered CKD diagnosis, with or without available pathological eGFR and/or UACR values the estimation fell to 3.7% (2.6–4.8) [7].

Finally, only a few studies like ours without biological results have used medico-administrative database that combined health claims with diagnosis and information on drugs, visits, or procedures to improve their algorithms. In the Lazio region, a study has combined different health information systems: the hospital discharge registry, the ticket exemption registry, the outpatient specialist service information system, the drug dispensing registry, the regional registry of causes of death, and the regional health assistance files [18]. The crude prevalence rate of CKD in the Lazio region was estimated at 1.76% (95%CI 1.75, 1.78). When applied to patients at Gemelli Hospital, an academic medical centre in Rome, the prevalence was estimated at 8.8% (95%CI 8.5–9.1) [38]. A recent French study used machine learning to identify patients with CKD based on a 1/97th representative sample of the general French population [39]. Their estimated prevalence was 0.8% for non-dialysis-dependent CKD.

Performances of all these algorithms are linked to the type of data used, diagnoses and procedural codes used in hospital-based database, and/or prescriptions of specific drugs or laboratory biological tests. The transportability of these studies from one country to another is also difficult due to the different types of data available. Indication bias due to the fact that laboratory tests and procedures are linked to a patient's characteristics must also be considered [40].

Number of individuals on RRT was estimated at 81 695 in 2021, lower than the 92 535 patients published by the French National REIN registry [41]. Missing cases are probably stable kidney transplant patients with no sufficient health claims to be detected by the RENAGLO-EPXERT algorithm.

Table 4: Assignment status with three other thresholds of eGFR level.

		eGFR CONSTANCES cohort							
RENALGO-EXPERT Algorithm		eGFR < 90		eGFR ≥ 90		eGFR < 45	eGFR ≥ 45	eGFR < 30	eGFR ≥ 30
CKD+		TP = 3 734 (1.9%)		FP = 5 081 (2.6%)		TP = 188 (0.1%)	FP = 8627 (4.4%)	TP = 56 (0.03%)	FP = 8759 (4.5%)
CKD−		FN = 64 682 (32.9%)		TN = 123 150 (62.6%)		FN = 75(0.04%)	TN = 187 757 (95.5%)	FN = 5(0%)	TN = 1987 827 (95.5%)
eGFR thresholds	**Se**	**Sp**	**PPV**	**NPV**	**LR+**		**LR−**	**Accuracy**	**Cohen's kappa coefficient**
eGFR < 90	5.4%	96.0%	42.4%	65.6%	1.35		0.99	64.5%	1.2% (0.9–1.5%)
eGFR < 45	71.5%	95.6%	2.1%	99.96%	16.25		0.3	95.6%	3.9% (3.3–4.4%)
eGFR < 30	91.8%	95.5%	0.6%	99.97%	20.4		0.09	95.5%	1.9% (1.6–2.1%)

TP, true positive; FN, false negative; FP, false positive; TN, true negative, eGFR in ml/min/1.73 m².

Performance of the RENALGO-EXPERT algorithm in the CONSTANCES cohort

Many algorithms used to identify CKD in hospital discharge databases have high specificity but relatively low sensitivity [13–16, 42]. Adding medication, i.e. CKD targeted drugs to the algorithm, by reducing false positives, improves specificity and reduces sensitivity. Sensitivity generally improves as the observation window gets longer because it reduces false negatives due to more information being taken into account.

The Lazio algorithm, like ours, showed better performance when applied to hospital patients, with a sensitivity of 51.0%, specificity of 96.5%, PPV of 64.5%, and NPV of 94.0% to detect CKD defined by eGFR <60 ml/min/1.73 m². This better performance (higher PPV and NPV) could be explained by the fact that it was applied to patients who had laboratory measurements prescribed during hospitalization, emergency room access, or ambulatory care that allows better defining CKD. When applied to the general population, the Lazio CKD prevalence was only 1.8% with lower performance due to low sensitivity.

Applied to a sample of the general population, our RENALGO-EXPERT algorithm had a low performance to detect CKD defined by eGFR <60 ml/min/1.73 m². In fact, by construction, the RENALGO-EXPERT algorithm designed around health claims data from a 1-year period, primarily identifies patients following established and recommended care pathways. Therefore, it may not readily detect undiagnosed individuals or those receiving inadequate care. The lack of specific drugs or procedures implies to use various combinations that makes describing the algorithm particularly complex (Supplementary Table S5). The previous validation of the RENAGLO-EPXERT algorithm in the French Childhood Cancer Survivor Study cohort showed better performance in a selected population who are likely to receive optimal care [20].

Considering the possible 'false positives', individuals that are identified by RENALGO-EXPERT are not at all uninteresting for a targeted screening strategy. Considering the possible 'false negatives', in France, each year around 60 million measurements of serum creatinine are reimbursed by the National Health Insurance (https://assurance-maladie.ameli.fr/etudes-et-donnees/actes-biologie-medicale-type-prescripteur-biolam). Since 2012, medical analysis laboratories are asked to report serum creatinine results associated, for the evaluation of kidney function, by an estimate of the GFR by the CKD-EPI equation (https://www.has-sante.fr/upload/docs/application/pdf/2012-10/evaluation_du_debit_de_filtration_glomerulaire_et_du_dosage_de_la_creatininemie_dans_le_diagnostic_de_la_maladie_renale_chronique_chez_ladulte_-_fiche_buts.pdf). However, especially at an early stage, this does not necessarily lead to a specific modification of the care that could have made it possible to identify the patients. Nevertheless, when focusing on a subgroup with a higher likelihood of being diagnosed, the algorithm's sensitivity significantly improves to 91.8% with a reference eGFR of 30 ml/min/1.73 m².

The main limitations of using CONSTANCES to study the RENALGO-EXPERT algorithm's performance was the relative low age of the volunteers and the small number of volunteers who had two creatinine measurements and an albuminuria test. Because chronicity criterion could not be used, we may have underestimated the sensitivity by including false positive cases. On the other hand, as in the paper of van Oosten and colleagues, sensitivity was higher in volunteers aged <50 years. This could be explained by the fact that young people are less likely to use healthcare and have fewer comorbidities such as diabetes and

Table 5: Performance of the RENALGO-EXPERT algorithm according to the level of eGFR among two subpopulations: 5716 volunteers with diabetes, 26 974 with hypertension, 110 687 < 50 years, 85 960 ≥ 50 years, 91 438 men, and 105 209 women.

	RENALGO-EXPERT Algorithm	eGFR CONSTANCES cohort	
		eGFR < 60	eGFR ≥ 60
Diabetes	CKD+	TP = 126 (2.2%)	FP = 1 115 (19.5%)
	CKD−	FN = 115 (2.0%)	TN = 4 360 (76.3%)
Hypertension	CKD+	TP = 427 (1.6%)	FP = 2 688 (9.9%)
	CKD−	FN = 753 (2.8%)	TN = 23 106 (85.7%)
<50 years	CKD+	TP = 58 (0.05%)	FP = 3 950 (3.6%)
	CKD−	FN = 101 (0.1%)	TN = 106 578 (96.3%)
≥50 years	CKD+	TP = 490 (0.6%)	FP = 4 317 (5.0%)
	CKD−	FN = 1 560 (1.8%)	TN = 79 593 (92.6%)
Men	CKD+	TP = 333 (0.4%)	FP = 3 057 (3.3%)
	CKD−	FN = 749 (0.8%)	TN = 87 299 (95.5%)
Women	CKD+	TP = 215 (0.2%)	FP = 5 210 (5.0%)
	CKD−	FN = 912 (0.9%)	TN = 98 872 (94.0%)

TP, true positive; FN, false negative; FP, false positive; TN, true negative; eGFR in ml/min/1.73 m²;

	Se (%)	Sp (%)	PPV (%)	NPV (%)	LR+ (%)	LR− (%)	Accuracy (%)	Cohen's kappa coefficient (%)
Diabetes	52.3	79.6	10.2	97.4	2.56	0.60	78.5	10.7 (08.3–13.10)
Hypertension	36.2	89.6	13.7	96.8	3.48	0.71	87.2	14.5 (12.8–16.1)
<50 years	36.5	96.4	1.4	99.9	10.14	0.66	96.3	2.5 (1.8–3.2)
≥50 years	23.9	94.8	10.2	98.1	4.6	0.8	93.2	11.3 (10.2–12.5)
Men	30.8	96.6	9.8	99.1	9.06	0.72	95.8	13.3 (11.9–14.8)
Women	19.1	95.0	4.0	99.1	3.82	0.85	94.2	4.9 (4.0–5.7)

Se, sensitivity; Sp, specificity; PPV, positive predictive value; NPV, negative predictive value.

hypertension, so they are less likely to be false positives. The CONSTANCES volunteers were under 70 years of age, which did not allow us to evaluate the performance of our algorithm in the elderly, a high-risk population for CKD. Analysis of misclassified cases has shown that our reference to define CKD with only one value of eGFR may be questionable.

Perspectives

Even if the use of the CONSTANCES cohort may be questionable due to its composition (i.e. the use of a single eGFR value and the low number of individuals with CKD), this validation phase showed us a poor ability of the algorithm to identify patients when they had an eGFR <60. This low sensitivity suggests the risk of underestimating prevalence. In subgroups with a higher risk of CKD, the performance of RENALGO-EXPERT improved a little. Because positive and negative predictive values do inherently vary with pre-test probability (e.g. changes in population disease prevalence), the performance of the RENALGO-EXPERT algorithm will be tested in further cohorts including high-risk patients.

In Table 6 advantages and limits of various methods to estimate CKD prevalence are presented. While the ideal situation would be to have two creatinine determinations and a proteinuria test in a representative sample of people and in each subgroup of interest, each territory must make do with its own available data.

Very few countries have easy access to all routine biological data. Until a database that centralizes all biological test results becomes available in France, the RENALGO-EXPERT algorithm can be used as a tool to monitor CKD prevalence within the French National Health Data System (SNDS) and therefore guide health policy planning.

Although the sensitivity of the RENALGO-EXPERT algorithm is low in the general population, its high specificity is interesting to identify high-risk groups of CKD. Studies that aim to evaluate care and trajectories of these patients could be carried out on such a sample more likely to be representative of the target population. This algorithm will now be used by French Health authorities as a contributing tool for thought on expenditure and the evolution of the care offering. However, because patients not diagnosed or without specific medical care are not detected, this monitoring will have to be associated by the development of additional tools.

Although databases differ widely from one country to another, due to reimbursement methods and social security coverage, standardizing such an algorithm would enable international comparisons.

In an attempt to improve our identification tool, an algorithm using Artificial Intelligence is under development, i.e. RENALGO-AI.

CONCLUSION

By virtue of its construction, the RENALGO-EXPERT algorithm can identify patients whose care pathway is close to the usual and specific CKD pathways. It does not identify patients who have not been diagnosed or whose care is inappropriate or at early stage with stable GFR. This is an inherent limitation of this kind of approach, which is based on healthcare consumption rather than biological assays on a representative sample. However, the estimated prevalence of CKD in France at 8%–10%

Table 6: Summary of various approaches to estimate CKD prevalence.

Sources	Information available	Advantages	Limits	consequence on CKD prevalence estimation	Examples (reference)
Cross-sectional studies with one measurement	serum creatinine, urine albumin	partial validated definition of CKD, possibility to identify asymptomatic individuals	Costly and time-consuming studies to set up, on sample basis (representation?), no chronicity criteria	questionable extrapolation, over estimation of the sensitivity	France: MONALISA USA: NHANES
Cross-sectional studies with two measurements	serum creatinine, urine albumin	validated definition of CKD, possibility to identify asymptomatic individuals	Costly and time-consuming studies to set up, on sample basis (representativity?)	questionable extrapolation, over estimation of the sensitivity	Morocco
Electronic medical records	serum creatinine, urine albumin	validated definition of CKD, chronicity, possibility to identify asymptomatic individuals	Depending on accessibility of such database, on sample basis (representativity?)	questionable extrapolation	Canada: Canadian Primary Care Sentinel Surveillance Network
Diagnosis at hospital discharge	diagnosis	classification done by physicians, data reuse	depending on accessibility of such database, quality of coding, selection bias (hospitalized patients)	great underestimation	Ontario study
Diagnosis at hospital discharge or outpatient claims	diagnosis	classification done by physicians, date reuse	depending on accessibility of such database, quality of coding, selection bias (people who sought medical care)	underestimation	USRDS
Diagnosis and laboratory results	diagnosis and serum creatinine +/− urine albumin	validated definition and classification done by physicians, date reuse	depending on accessibility of such database, quality of coding, selection bias (people who sought medical care)	underestimation but probably close to the best possible estimate	CaReMe CKD study SCREAM project
Medico-administrative databases, experts, algorithm	diagnosis, drugs, visits, procedures	data reuse	depending on accessibility of such database and quality of the algorithm, selection bias (people who sought medical care)	hard to estimate	Italy: Lazio study France: RENALGO-EXPERT
Medico-administrative databases, algorithm from Artificial Intelligence	diagnosis, drugs, visits, procedures	data reuse	depending on accessibility of such database and quality of the algorithm, algorithm interpretability, selection bias (people who sought medical care)	hard to estimate	France

is close to that expected. As suggested by some recent publication, weighting to account for individuals less regularly monitored may provide more reliable prevalence estimates [40].

SUPPLEMENTARY DATA

Supplementary data are available at *ckj* online.

ACKNOWLEDGEMENTS

The authors thank the team of the 'Population-based cohorts unit' (Cohortes en population) that designed and manages the CONSTANCES cohort study. They also thank the French National Health insurance fund ('Caisse nationale d'assurance maladie', Cnam) and its health screening centres ('Centres d'examens de santé'), which are collecting a large part of the data, as well as the French national old-age insurance fund (Caisse nationale d'assurance vieillesse, Cnav) for its contribution to the constitution of the cohort, and ClinSearch, Asqualab, and Eurocell, which are conducting data quality control.

FUNDING

The CONSTANCES cohort study was supported and funded by the French National Health insurance fund (Caisse nationale d'assurance maladie, Cnam). CONSTANCES is a national infrastructure for biology and health ('Infrastructure nationale en biologie et santé') and benefits from a grant from the French national agency for research (ANR-11-INBS-0002). CONSTANCES is also partly funded to a small extent by industrial companies, notably in the healthcare sector, within the framework of Public-Private Partnerships. None of these funding sources had any role in the design of the study, collection, and analysis of data or decision to publish.

AUTHORS' CONTRIBUTIONS

The authors confirm contribution to the paper as follows: RENALGO-EXPERT development: C.C., M.R., M.L., M.O., E.M., L.H., P.C., G.E.F., H.A., M.M., M.I., and B.M.S.; study conception and design: C.C., M.R., B.M.S., and S.K.; data collection: G.M., Z.M., and K.S.; analysis and interpretation of results: C.C., M.R., D.Z., and S.K.; draft manuscript preparation: C.C., M.R., M.O., E.M., L.H., P.C., G.E.F., H.A., M.M., B.M.S., and K.S.Z. All authors reviewed the results and approved the final version of the manuscript.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

Due to legal restrictions SNDS or CONSTANCES data cannot be made publicly available. More information regarding data access can be found at https://www.constances.fr/guide-acces-donnees.pdf and https://www.health-data-hub.fr/cesrees.

REFERENCES

1. Bikbov B, Purcell CA, Levey AS *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 North Am Ed* 2020;**395**:709–33. https://doi.org/10.1016/S0140-6736(20)30045-3
2. Huijben JA, Kramer A, Kerschbaum J *et al*. Increasing numbers and improved overall survival of patients on kidney replacement therapy over the last decade in Europe: an ERA Registry study. *Nephrol Dial Transplant* 2023;**38**:1027–40. https://doi.org/10.1093/ndt/gfac165
3. Caillet A, Mazoué F, Wurtz B *et al*. Which data in the French registry for advanced chronic kidney disease for public health and patient care? *Nephrol Ther* 2022;**18**:228–36. https://doi.org/10.1016/j.nephro.2022.01.004
4. Couchoud C, Stengel B, Landais P *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.
5. Issad B, Galland R, Merle V *et al*. Prévalence de l'IRCT et part des différentes modalités de traitement. *Néphrologie Thérapeutique* 2022;**18**:18/5S–e15–e18/5S–e20. https://doi.org/10.1016/S1769-7255(22)00563-6
6. Bello AK, Levin A, Manns BJ *et al*. Effective CKD care in European countries: challenges and opportunities for health policy. *Am J Kidney Dis Off J Natl Kidney Found* 2015;**65**:15–25. https://doi.org/10.1053/j.ajkd.2014.07.033
7. Sundström J, Bodegard J, Bollmann A *et al*. Prevalence, outcomes, and cost of chronic kidney disease in a contemporary population of 2·4 million patients from 11 countries: the CaReMe CKD study. *Lancet Reg Health—Eur* 2022;**20**:100438. https://doi.org/10.1016/j.lanepe.2022.100438
8. Alwan A, Maclean DR, Riley LM *et al*. Monitoring and surveillance of chronic non-communicable diseases: progress and capacity in high-burden countries. *Lancet Lond Engl* 2010;**376**:1861–8. https://doi.org/10.1016/S0140-6736(10)61853-3
9. KDIGO. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney Int* 2013;**Suppl**:1–150.
10. Writing Group for the CKD Prognosis Consortium, Grams ME, Coresh J, Matsushita K *et al*. Estimated glomerular filtration rate, albuminuria, and adverse outcomes: an individual-participant data meta-analysis. *JAMA* 2023;**330**:1266–77.
11. van Oosten MJM, Brohet RM, Logtenberg SJJ *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. https://doi.org/10.1093/ckj/sfaa167
12. Fleet JL, Dixon SN, Shariff SZ *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. https://doi.org/10.1186/1471-2369-14-81
13. Ronksley PE, Tonelli M, Quan H *et al*. Validating a case definition for chronic kidney disease using administrative data. *Nephrol Dial Transplant* 2012;**27**:1826–31. https://doi.org/10.1093/ndt/gfr598
14. Muntner P, Gutiérrez OM, Zhao H *et al*. Validation study of Medicare claims to identify older US adults with CKD using the reasons for geographic and racial differences in stroke (REGARDS) study. *Am J Kidney Dis* 2015;**65**:249–58. https://doi.org/10.1053/j.ajkd.2014.07.012
15. Grams ME, Plantinga LC, Hedgeman E *et al*. Validation of CKD and related conditions in existing data sets: a systematic review. *Am J Kidney Dis* 2011;**57**:44–54. https://doi.org/10.1053/j.ajkd.2010.05.013

16. Vlasschaert MEO, Bejaimal SAD, Hackam DG *et al*. Validity of administrative database coding for kidney disease: a systematic review. *Am J Kidney Dis* 2011;**57**:29–43. https://doi.org/10.1053/j.ajkd.2010.08.031
17. Bello A, Hemmelgarn B, Manns B *et al*. for Alberta Kidney Disease Network. Use of administrative databases for healthcare planning in CKD. *Nephrol Dial Transplant* 2012;**27**:iii12–8. https://doi.org/10.1093/ndt/gfs163
18. Marino C, Ferraro PM, Bargagli M *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. https://doi.org/10.1186/s12882-020-1689-z
19. Tuppin 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–S67. https://doi.org/10.1016/j.respe.2017.05.004
20. Mansouri I, Raffray M, Lassalle M *et al*. An algorithm for identifying chronic kidney disease in the French national health insurance claims database. *Nephrol Ther* 2022;**18**:255–62. https://doi.org/10.1016/j.nephro.2022.03.003
21. Moulis G, Lapeyre-Mestre M, Palmaro A *et al*. French health insurance databases: what interest for medical research? *Rev Médecine Interne* 2015;**36**:411–7. https://doi.org/10.1016/j.revmed.2014.11.009
22. Tuppin P, de RL, Weill A *et al*. French national health insurance information system and the permanent beneficiaries sample. *Rev Epidemiol Sante Publique* 2010;**58**:286–90.
23. 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**:954–62. https://doi.org/10.1002/pds.4233
24. CONSTANCES Team, Zins M, Goldberg M. The French CONSTANCES population-based cohort: design, inclusion and follow-up. *Eur J Epidemiol* 2015;**30**:1317–28. https://doi.org/10.1007/s10654-015-0096-4
25. Delanaye P, Glassock RJ, De Broe ME. Epidemiology of chronic kidney disease: think (at least) twice! *Clin Kidney J* 2017;**10**:370–4. https://doi.org/10.1093/ckj/sfw154
26. Glassock RJ, Warnock DG, Delanaye P. The global burden of chronic kidney disease: estimates, variability and pitfalls. *Nat Rev Nephrol* 2017;**13**:104–14. https://doi.org/10.1038/nrneph.2016.163
27. De Broe ME, Gharbi MB, Zamd M *et al*. Why overestimate or underestimate chronic kidney disease when correct estimation is possible? *Nephrol Dial Transplant* 2017;**32**:ii136–41. https://doi.org/10.1093/ndt/gfw267
28. Venuthurupalli SK, Hoy WE, Healy HG *et al*. CKD screening and surveillance in Australia: past, present, and future. *Kidney Int Rep* 2018;**3**:36–46. https://doi.org/10.1016/j.ekir.2017.09.012
29. Carrero JJ, Elinder CG. The Stockholm CREAtinine Measurements (SCREAM) project: fostering improvements in chronic kidney disease care. *J Intern Med* 2022;**291**:254–68. https://doi.org/10.1111/joim.13418
30. Bongard V, Dallongeville J, Arveiler D *et al*. Estimation et caractérisation de l'insuffisance rénale chronique en France. *Ann Cardiol Angéiologie* 2012;**61**:239–44. https://doi.org/10.1016/j.ancard.2012.03.003
31. Bruck K, Stel VS, Gambaro G *et al*. CKD prevalence varies across the European general population. *J Am Soc Nephrol* 2016;**27**:2135–47.
32. Kibria GMA, Crispen R. Prevalence and trends of chronic kidney disease and its risk factors among US adults: an analysis of NHANES 2003-18. *Prev Med Rep* 2020;**20**:101193. https://doi.org/10.1016/j.pmedr.2020.101193
33. Zhang L, Wang F, Wang L *et al*. Prevalence of chronic kidney disease in China: a cross-sectional survey. *Lancet North Am Ed* 2012;**379**:815–22. https://doi.org/10.1016/S0140-6736(12)60033-6
34. Benghanem Gharbi M, Elseviers M, Zamd M *et al*. Chronic kidney disease, hypertension, diabetes, and obesity in the adult population of Morocco: how to avoid "over"- and "under"-diagnosis of CKD. *Kidney Int* 2016;**89**:1363–71. https://doi.org/10.1016/j.kint.2016.02.019
35. Bello AK, Ronksley PE, Tangri N *et al*. Prevalence and demographics of CKD in Canadian primary care practices: a cross-sectional study. *Kidney Int Rep* 2019;**4**:561–70. https://doi.org/10.1016/j.ekir.2019.01.005
36. Hill NR, Fatoba ST, Oke JL *et al*. Global Prevalence of chronic kidney disease—a systematic review and meta-analysis. Remuzzi G, editor. *PLoS One* 2016;**11**:e0158765. https://doi.org/10.1371/journal.pone.0158765
37. Collins AJ, Chen SC, Gilbertson DT *et al*. CKD surveillance using administrative data: impact on the health care system. *Am J Kidney Dis* 2009;**53**:S27–S36. https://doi.org/10.1053/j.ajkd.2008.07.055
38. Ferraro PM, Agabiti N, Angelici L *et al*. Validation of a classification algorithm for chronic kidney disease based on health information systems. *J Clin Med* 2022;**11**:2711. https://doi.org/10.3390/jcm11102711
39. Dardim K, Fernandes J, Panes A *et al*. Incidence, prevalence, and treatment of anemia of non-dialysis-dependent chronic kidney disease: a retrospective database study in France. Bennett K, editor. *PLoS ONE* 2023;**18**:e0287859. https://doi.org/10.1371/journal.pone.0287859
40. Mazhar F, Sjölander A, Fu EL *et al*. Estimating the prevalence of chronic kidney disease while accounting for nonrandom testing with inverse probability weighting. *Kidney Int* 2023;**103**:416–20. https://doi.org/10.1016/j.kint.2022.10.027
41. Couchoud C, Lassalle M. *REIN Annual Report 2021. Agence De la Biomédecine*, France; https://www.agence-biomedecine.fr/IMG/pdf/rapport_rein_2021_2023-06-26.pdf (13 May 2024, date last accessed)
42. van Oosten MJM, Logtenberg SJJ, Edens MA *et al*. Health claims databases used for kidney research around the world. *Clin Kidney J* 2021;**14**:84–97. https://doi.org/10.1093/ckj/sfaa076

Received: 19.12.2023; Editorial decision: 16.3.2024