ORIGINAL ARTICLE

The validity of Dutch health claims data for identifying patients with chronic kidney disease: a hospital-based study in the Netherlands

Manon J.M. van Oosten[1], Richard M. Brohet[2], Susan J.J. Logtenberg[3], Anneke Kramer [1], Lambert D. Dikkeschei[4], Marc H. Hemmelder[5,6], Henk J.G. Bilo[2,7,8], Kitty J. Jager[1] and Vianda S. Stel[1]

[1]Department of Medical Informatics, Amsterdam UMC, Amsterdam Public Health Research Institute, University of Amsterdam, Amsterdam, The Netherlands, [2]Diabetes Research Center and Department of Epidemiology and Statistics, Isala Hospital, Zwolle, The Netherlands, [3]Department of Internal Medicine, Diakonessenhuis, Utrecht, The Netherlands, [4]Department of Clinical Chemistry, Isala Hospital, Zwolle, The Netherlands, [5]Dutch Renal Registry (Renine), Nefrovisie Foundation, Utrecht, The Netherlands, [6]Department of Internal Medicine, Medical Center Leeuwarden, Leeuwarden, The Netherlands, [7]Department of Internal Medicine, University Medical Center, Groningen, The Netherlands and [8]Faculty of Medicine, Groningen University, Groningen, The Netherlands

Correspondence to: Manon J.M. van Oosten; E-mail: m.j.m.vanoosten@amsterdamumc.nl

ABSTRACT

Background. Health claims data may be an efficient and easily accessible source to study chronic kidney disease (CKD) prevalence in a nationwide population. Our aim was to study Dutch claims data for their ability to identify CKD patients in different subgroups.

Methods. From a laboratory database, we selected 24 895 adults with at least one creatinine measurement in 2014 ordered at an outpatient clinic. Of these, 15 805 had ≥ 2 creatinine measurements at least 3 months apart and could be assessed for the chronicity criterion. We estimated the validity of a claim-based diagnosis of CKD and advanced CKD. The estimated glomerular filtration rate (eGFR)-based definitions for CKD (eGFR < 60 mL/min/1.73 m^2) and advanced CKD (eGFR < 30 mL/min/1.73 m^2) satisfying and not satisfying the chronicity criterion served as reference group. Analyses were stratified by age and sex.

Results. In general, sensitivity of claims data was highest in the population with the chronicity criterion as reference group. Sensitivity was higher in advanced CKD patients than in CKD patients {51% [95% confidence interval (CI) 47–56%] versus 27% [95% CI 25–28%]}. Furthermore, sensitivity was higher in young versus elderly patients. In patients with advanced CKD, sensitivity was 72% (95% CI 62–83%) for patients aged 20–59 years and 43% (95% CI 38–49%) in patients ≥ 75 years. The specificity of CKD and advanced CKD was $\geq 99\%$. Positive predictive values ranged from 72% to 99% and negative predictive values ranged from 40% to 100%.

Conclusion. When using health claims data for the estimation of CKD prevalence, it is important to take into account the characteristics of the population at hand. The younger the subjects and the more advanced the stage of CKD the higher the sensitivity of such data. Understanding which patients are selected using health claims data is crucial for a correct interpretation of study results.

Keywords: CKD, GFR, health claims data, health claims database, validity study

INTRODUCTION

In recent decades, health insurance claims data have become available as a source of big data. Health claims databases often contain already well-defined data sets and hold information on patient demographics and healthcare resource use in a non-experimental setting over large populations. It has been suggested that health claims databases may have considerable advantages in calculating disease prevalence over large populations and observing trends over longer periods of time [1, 2].

Typically, health claims data lack both clinical and laboratory data and the identification of patients with specific diseases is solely based on specific diagnosis codes. This entails an inherent danger of inaccurate identification and possible under-coding or overcoding of diagnoses [3]. Validity studies are necessary to investigate whether health claims data can provide reliable estimates of the frequency of these diseases.

Only a few studies have assessed the accuracy of health claims data in identifying patients with chronic kidney disease (CKD) not treated with renal replacement therapy [4–9]. These studies provided limited information on the validity of specific patient subgroups. Understanding the relationship between patient characteristics and the ability to identify them with health claims data may assist in assessing the value of health claims in estimating CKD prevalence in those subgroups.

Therefore, our study aims to determine the validity of Dutch health claims data in identifying CKD patients in various patient subgroups (defined by age and sex) and for different definitions of CKD, using a hospital-based database in the Netherlands.

MATERIALS AND METHODS

Study population

Serum creatinine measurements from a regional medical laboratory serving general practitioners (GPs) and a hospital in the city of Zwolle, the Netherlands, served as a reference. There were no other large medical laboratories in this region. From this laboratory database, we selected adults (≥18 years) with at least one serum creatinine measurement between 1 January and 31 December 2014. Information in the laboratory database included the patient's date of birth and sex, the value and the measurement date of serum creatinine, the type of physician ordering the measurement (GP or medical specialist) and the care setting (primary care; secondary care divided in outpatients versus inpatients). Data on these individuals were linked to the health claims database of the Zwolle hospital which includes claims data of all delivered hospital care for a specific medical condition or complaint. This medical care can be delivered during a hospital admittance or during (a) visit(s) at the outpatient clinic. Patients treated with dialysis or kidney transplantation were identified using health claims data and excluded from our study [10].

For our main analyses, we selected an outpatient population in which the last serum creatinine measurement was ordered in the outpatient clinic. We consider this the best proxy for the general population, as during hospitalization kidney function can temporarily deteriorate without the patient having CKD and because CKD patients solely known to a GP cannot be detected with hospital claims. Secondary analyses were performed for a GP and inpatient population, in which the last serum creatinine measurement was ordered by a GP or in an inpatient setting.

Identification of CKD patients

In the Netherlands, hospital care is reimbursed via physician claims named diagnosis treatment combinations (DBCs), a system similar to Diagnosis Procedure Codes. Every hospital DBC code corresponds to a specific medical condition in a specific medical discipline [11]. This DBC comprises all delivered hospital care for this condition, for example, care delivered during a hospital admittance or at an outpatient clinic as well as laboratory or radiology procedures. Table 1 provides an overview of the identification methods of CKD patients using the health claims and laboratory databases.

Hospital health claims database. Patients with a DBC code 0313.11.324 'chronic renal insufficiency eGFR 30–60 mL/min/1.73 m^2' and/or a DBC code 0313.11.325 'chronic renal insufficiency eGFR <30 mL/min/1.73 m^2' were defined as patients with a claim-based diagnosis of CKD or advanced CKD, respectively.

Laboratory database. Kidney function was estimated by calculating the estimated glomerular filtration rate (eGFR) for each creatinine measurement in 2014 using the Chronic Kidney Disease Epidemiology Collaboration formula. Ethnicity status was not included in the eGFR equation because this was not available. For the diagnosis of CKD (Stages 3–5) and advanced CKD (Stages 4–5), we used four different definitions based on a single creatinine measurement or ≥2 measurements at least 3 months apart, thereby satisfying the chronicity criterion according to international guidelines (Table 1) [12]. In cases where different creatinine measurements of a patient resulted in different CKD classification (i.e. no CKD, CKD or advanced CKD), we classified this person in the category with the highest eGFR to ensure that a temporary decrease in eGFR did not result in a premature diagnosis of chronic (advanced) CKD.

Statistical analysis

We estimated the validity of the claim-based diagnoses of CKD and advanced CKD using the four eGFR-based CKD definitions applied to the laboratory database as the reference group (see Table 1). Stratified analyses were performed by sex and age groups (i.e. 20–59 years, 60–74 years, ≥75 years). Since the sensitivity of claims data was relatively low in patients ≥75 years of age, we performed a subgroup analysis with patients under the age of 75 years. For our main analysis, we used eGFR

Table 1. Identification of CKD patients using a health claims database and a laboratory database

Health claims database		Laboratory database	
Study population	Claim-based diagnosis	Reference group	eGFR-based definition[a]
CKD	DBC code 0313.11.324: eGFR 30–60 mL/min/1.73 m^2 or DBC code 0313.11.325: eGFR <30 mL/min/1.73 m^2	CKD$_{single}$	One eGFR calculation of <60 mL/min/1.73 m^2 (not satisfying chronicity criterion)
		CKD$_{chron}$	≥2 eGFR calculations of <60 mL/min/1.73 m at least 3 months apart (satisfying chronicity criterion)
Advanced CKD	DBC code 0313.11.325: eGFR <30 mL/min/1.73 m^2	Advanced CKD$_{single}$	One eGFR calculation of <30 mL/min/1.73 m^2 (not satisfying chronicity criterion)
		Advanced CKD$_{chron}$	≥2 eGFR calculations of <30 mL/min/1.73 m^2 at least 3 months apart (satisfying chronicity criterion)

[a]For calculation of the eGFR the Chronic Kidney Disease Epidemiology Collaboration formula was used.
CKD, chronic kidney disease; DBC, 'diagnosis treatment combination', in Dutch: 'diagnose behandeling combincatie'; eGFR, estimated glomerular filtration rate.

calculations derived from creatinine measurements in an outpatient setting. Secondary analyses were performed for eGFR calculations conducted in GP and inpatient settings. We estimated the validity of health claims data by calculating the sensitivity (true-positive rate; the proportion of actual CKD patients correctly identified as such with health claims data), the specificity (true-negative rate; the proportion of actual negatives using the claim-based definition correctly identified as having no CKD using the eGFR-based definition), the positive predictive value (PPV; the probability that CKD is actually present among those with a claim-based diagnosis of CKD) and the negative predictive value (NPV; the probability that CKD is actually absent among those without a claim-based diagnosis of CKD) by using the eGFR-based CKD study populations as the reference group (see Supplementary data, Appendix 1). The CKD prevalence was calculated using the number of CKD patients identified using the eGFR-based definition of CKD divided by the total general population of Zwolle. In a separate analysis, CKD prevalence estimates were adjusted for age and sex using the Dutch general population of 2014 as a reference. Adjusted CKD prevalence was derived by applying the weights of the reference population to the observed variable specific prevalence (e.g. CKD prevalence per age group) in the Zwolle population. This weighted average provides a single summary CKD prevalence that would be expected if the region of Zwolle had the age and sex distribution of the reference population. SPSS 24.0 and SAS 9.4 were used for all calculations.

RESULTS

Baseline characteristics

We identified 67 773 individuals with at least one serum creatinine measurement in 2014 (Table 2). Their mean age was 60.5 (SD 16.9) years, 46% were male and the prevalence of CKD (eGFR <60 mL/min/1.73 m^2), based on a single creatinine measurement (CKD$_{single}$), was 19.1%, with 2.1% having an eGFR <30 mL/min/1.73 m^2. A subset of 36 504 individuals had ≥2 creatinine measurements in 2014 at least 3 months apart and could be assessed for satisfying the chronicity criterion. In this group, with a mean age of 63.8 (SD 15.6) years and 47% males, 20.8% of

individuals had an eGFR <60 mL/min/1.73 m^2 and 2.2% had an eGFR <30 mL/min/1.73 m^2.

In 24 895 outpatient individuals, 19.8% [95% confidence interval (CI) 19.3–20.3%] had an eGFR <60 mL/min/1.73 m^2 and 2.9% (95% CI 2.7–3.1%) an eGFR <30 mL/min/1.73 m^2 (Table 2). Of this outpatient population, 15 805 individuals had ≥2 creatinine measurements at least 3 months apart. Using the chronicity criterion 21.5% (95% CI 20.9–22.1%) had an eGFR <60 mL/min/1.73 m^2 and 3.4% (95% CI 3.1–3.7%) an eGFR <30 mL/min/1.73 m^2. The CKD prevalence adjusted for age and sex was lower compared with the unadjusted CKD prevalence in the outpatient study group (Table 2). In the same group of outpatient individuals, the unadjusted prevalence of CKD based on health claims was 4.1% (95% CI 3.9–4.4%) and 6.1% (95% CI 5.7–6.5%) in the population where the chronicity criteria could be taken into account. After adjustment for age and sex the prevalence was lower with, respectively, 2.9% (95% CI 2.7–3.1%) and 4.0% (95% CI 3.7–4.3%). The prevalence of advanced CKD (based on health claims data) was 1.5% (95% CI 1.3–1.6%) and 2.2% (95% CI 2.0–2.4%) in the population eligible to check for the chronicity criteria. After adjustment, the prevalence was lower with, respectively, 1.0% (95% CI 0.9–1.2%) and 1.5% (95% CI 1.3–1.7%).

Sensitivity

Total. Figure 1 presents the sensitivity of the claim-based diagnoses of CKD and advanced CKD. Sensitivity of the claim-based diagnosis of CKD was 20% when using CKD$_{single}$ as reference group. This means that 20% of the patients with an eGFR <60 mL/min/1.73 m^2 could be traced to have a CKD-related health claim (Figure 1). Sensitivity of CKD was 27% when the chronicity criterion was taken into account (CKD$_{chron}$). In patients with advanced CKD, sensitivity was 42% when using advanced CKD$_{single}$ as reference group, and 51% when using advanced CKD$_{chron}$ as reference group.

In general, the sensitivity of health claims data was higher in patients with advanced CKD as opposed to those with CKD. In addition, the sensitivity of health claims data was always higher when using eGFR-based diagnoses satisfying the chronicity criterion as the reference group.

Table 2. Baseline characteristics

Characteristics	Overall population[a]		Outpatient population[b]	
	Patients with ≥1 serum creatinine measurement	Patients allowing to test for chronicity[c]	Patients with ≥1 serum creatinine measurement	Patients allowing to test for chronicity[c]
Total				
n	67 773	36 504	24 895	15 805
Age (years), mean (SD)	60.5 (16.9)	63.8 (15.6)	58.1 (16.6)	60.5 (15.8)
Gender				
Male, n (%)	31 438 (46.4)	17 030 (46.7)	11 976 (48.1)	7795 (49.3)
Age groups (years), n (%)				
20–59	28 911 (42.7)	12 571 (34.4)	12 056 (48.4)	6762 (42.8)
60–74	23 817 (35.1)	13 993 (38.3)	8622 (34.6)	5853 (37.0)
≥75	15 045 (22.2)	9940 (27.2)	4217 (16.9)	3190 (20.2)
CKD				
n	12 978	7587	4924	3405
Unadjusted prevalence (95% CI)[d]	19.1 (18.8–19.4)	20.8 (20.4–21.2)	19.8 (19.3–20.3)	21.5 (20.9–22.1)
Adjusted prevalence (95% CI)[e]	10.6 (10.4–10.8)	10.3 (10.0–10.6)	13.1 (12.7–13.5)	13.0 (12.5–13.5)
Advanced CKD				
n	1443	820	719	536
Unadjusted prevalence (95% CI)	2.1 (1.9–2.2)	2.2 (2.0–2.4)	2.9 (2.7–3.1)	3.4 (3.1–3.7)
Adjusted prevalence (95% CI)	1.2 (1.1–1.3)	1.2 (1.1–1.3)	2.0 (1.8–2.2)	2.2 (2.0–2.4)

[a]The overall population patient population includes all patients with a serum creatinine measurement in the laboratory database regardless of the type of physician ordering the measurement and the clinical setting.
[b]The outpatient population includes all patients with their last serum creatinine measurement ordered in the outpatient clinic.
[c]The patient population allowing to test for chronicity are all patients with ≥2 serum creatinine measurements at least 3 months apart.
[d]The prevalence is based on the number of CKD patients identified using the eGFR-based definition of CKD.
[e]Adjusted for the age and sex distribution in the total Dutch population.

By age group and sex. Sensitivity was highest for patients aged 20–59 years and lowest in those ≥75 years of age, for all eGFR-based CKD definitions as reference group. In young patients with advanced CKD, sensitivity was 72% when using advanced CKD$_{chron}$ as reference group (Figure 1). Overall, the sensitivity was higher in men than in women (Figure 1). In contrast, in patients with advanced CKD below the age of 75 years, sensitivity was higher in women than in men. Of note, young female patients (20–59 years) with advanced CKD were most accurately identified with a sensitivity of 76% using CKD$_{chron}$ as reference group.

Age <75 years. Since the sensitivity of claims data was relatively low in patients ≥75 years of age, we performed a subgroup analysis with patients under the age of 75 years. As a result, the sensitivity increased, for example, in advanced CKD the sensitivity increased from 51% (Figure 1) to 62% when using advanced CKD$_{chron}$ as reference group (Figure 2).

Specificity, PPV and NPV

Overall and in all subgroups based on age and sex, specificity of CKD and advanced CKD was 99% or higher (Supplementary data, Table S1). PPVs ranged from 72% to 99% and NPVs ranged from 40% to 100%.

Age <75 years. Specificity, PPV and NPV of the subgroup of patients <75 years of age were comparable and are presented in the appendix (Supplementary data, Table S2).

Nephrological care

The majority of CKD patients without a concordant CKD health claim received adequate nephrological care (60%) (defined as having health claims related to CKD, nephrology or diabetes care). In CKD patients under the age of 75 years, this was even >90% (Supplementary data, Table S3).

GP and inpatient population

The baseline characteristics of the GP and inpatient study populations are described in Supplementary data, Table S4. In Supplementary data, Figure S1 and Tables S5–S7, we also present the results of the overall, the GP and inpatient study populations.

DISCUSSION

This study describes the validity of Dutch health claims data for the estimation of CKD prevalence, overall and in patient subgroups, in a hospital-based study. Since this study primarily assesses the value of health claims when estimating CKD prevalence in different patient subgroups, we mainly focus on the sensitivity. The 'overall' sensitivity of health claims data for the identification of CKD patients using the chronicity criterion as the reference group was 27%. The sensitivity of health claims data increased to 51% for patients with advanced CKD. Sensitivity of the claim-based diagnoses of CKD was substantially higher in young patients (age 20–59 years) and in men. A maximum of 76% was reached in young women with advanced
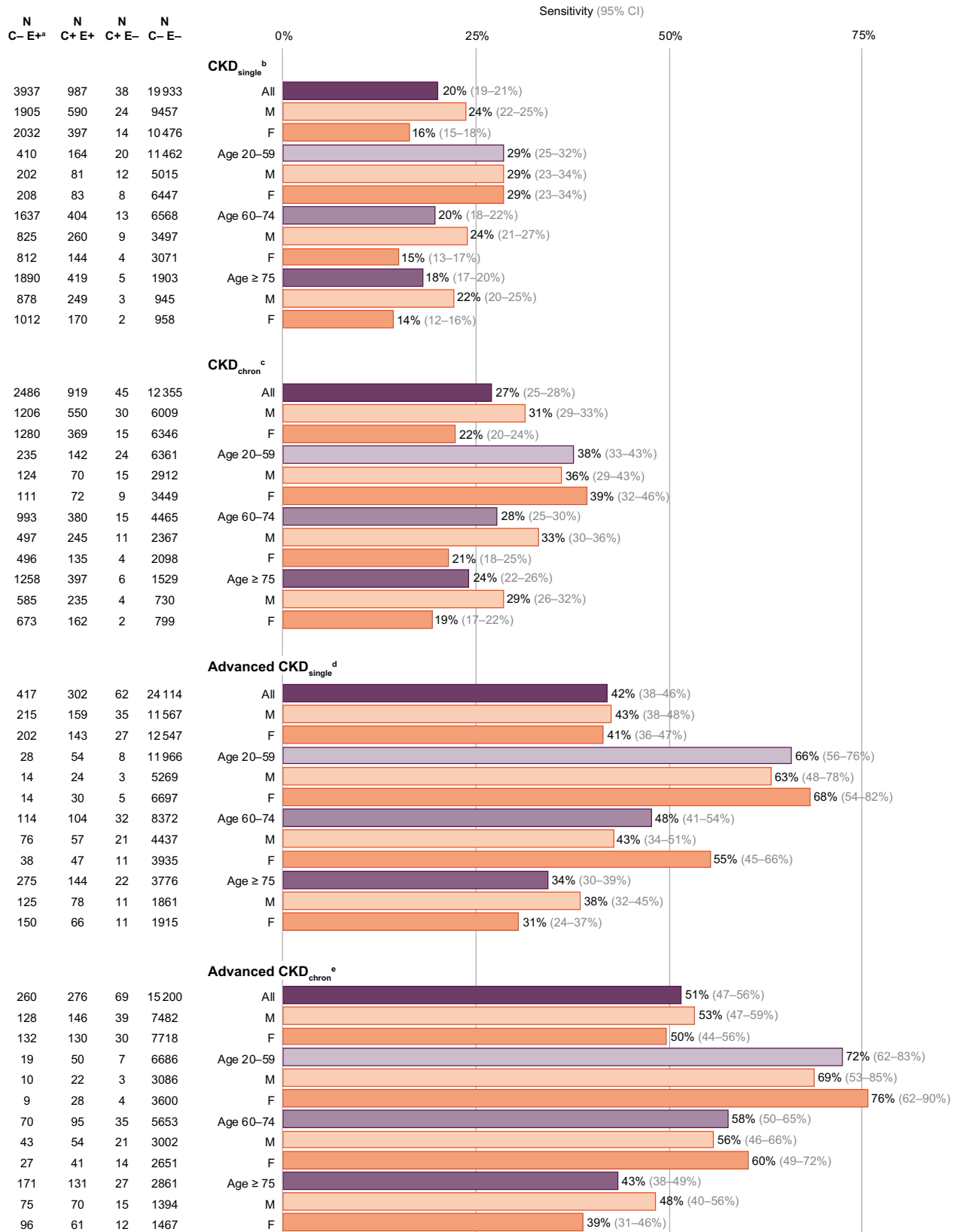
FIGURE 1: Sensitivity of claim-based diagnosis of CKD and of advanced CKD using four eGFR-based CKD definitions as the reference group, by age group and sex. F, female; M, male. [a]C, claim-based CKD diagnosis; E, eGFR-based CKD diagnosis. [b]CKD$_{single}$, one eGFR calculation <60 mL/min/1.73 m^2. [c]CKD$_{chron}$, ≥2 eGFR calculations <60 mL/min/1.73 m^2 at least 3 months apart. [d]Advanced CKD$_{single}$, one eGFR calculation <30 mL/min/1.73 m^2. [e]Advanced CKD$_{chron}$, ≥2 eGFR calculations <30 mL/min/1.73 m^2 at least 3 months apart.

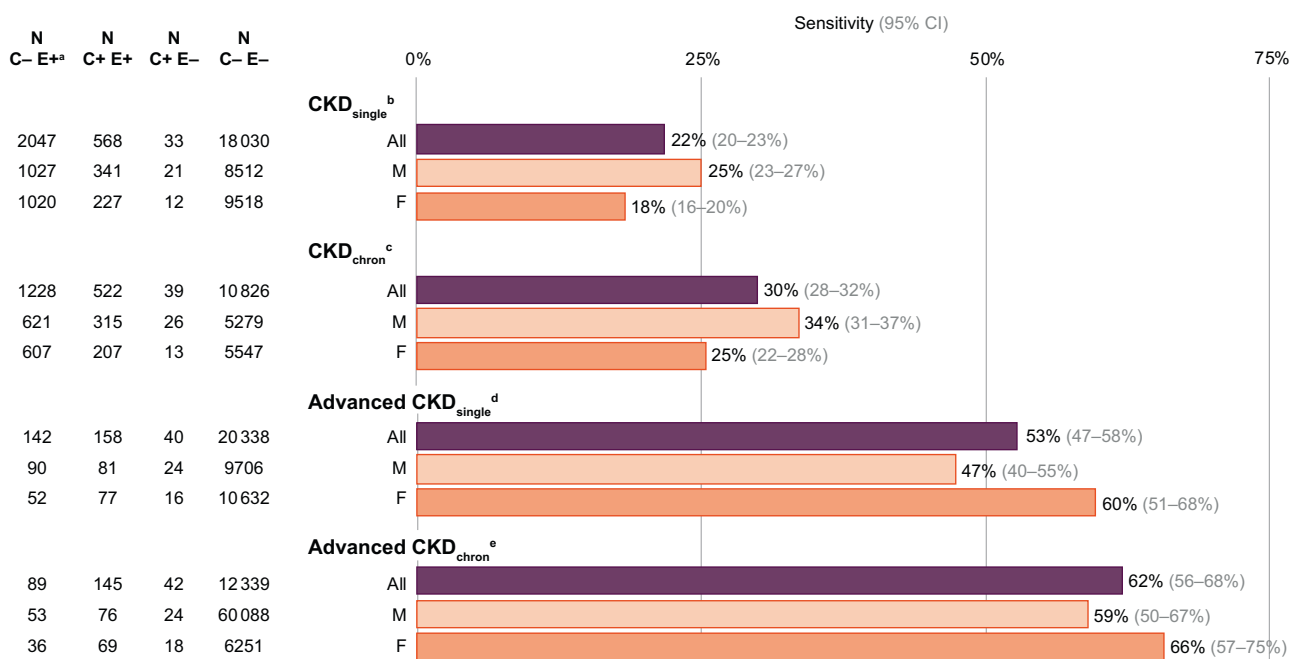| N C– E+[a]	N C+ E+	N C+ E–	N C– E–		
				CKD$_{single}$[b]	
2047	568	33	18030	All	22% (20–23%)
1027	341	21	8512	M	25% (23–27%)
1020	227	12	9518	F	18% (16–20%)
				CKD$_{chron}$[c]	
1228	522	39	10826	All	30% (28–32%)
621	315	26	5279	M	34% (31–37%)
607	207	13	5547	F	25% (22–28%)
				Advanced CKD$_{single}$[d]	
142	158	40	20338	All	53% (47–58%)
90	81	24	9706	M	47% (40–55%)
52	77	16	10632	F	60% (51–68%)
				Advanced CKD$_{chron}$[e]	
89	145	42	12339	All	62% (56–68%)
53	76	24	60088	M	59% (50–67%)
36	69	18	6251	F	66% (57–75%)

FIGURE 2: Sensitivity of claim-based diagnosis of CKD and of advanced CKD using four eGFR-based CKD definitions as the reference group, in patients <75 years. F, female; M, male. [a]C, claim-based CKD diagnosis, E, eGFR-based CKD diagnosis. [b]CKD$_{single}$, one eGFR calculation <60 mL/min/1.73 m^2. [c]CKD$_{chron}$, ≥2 eGFR calculations <60 mL/min/1.73 m^2 at least 3 months apart. [d]Advanced CKD$_{single}$, one eGFR calculation <30 mL/min/1.73 m^2. [e]Advanced CKD$_{chron}$, ≥2 eGFR calculations <30 mL/min/1.73 m^2 at least 3 months apart.

CKD. The specificity of CKD and advanced CKD was consistently high, whereas the PPV and NPV varied between the patient subgroups.

Sensitivity of health claims data in the estimation of CKD prevalence

Our study is the first describing validity of claims data in a European healthcare system for the identification of CKD patients. So far, four studies in Canada and the USA have assessed the validity of health claims data in identifying patients with CKD by comparing estimates of claim-based CKD prevalence with an eGFR-based CKD prevalence as reference group [4–7]. All studies were able to validate a claim-based diagnosis of CKD while two were additionally able to validate a claim-based diagnosis of advanced CKD [4, 5]. Only in one study, the eGFR-based CKD definition was based on ≥2 eGFR calculations, making it possible to take the chronicity criterion into account [4].

In line with our results, these studies concluded that health claims data have low sensitivity and high specificity for the identification of CKD patients [4–7]. The sensitivity for the identification of CKD patients ranged from 2.7% to 19.4% in patients with CKD and from 56.0% to 58.8% in patients with advanced CKD. The accuracy of health claims data in identifying CKD Stages 3–5 is slightly higher in our study using the chronicity criterion (sensitivity 27%), while slightly lower for advanced CKD using the chronicity criterion (in our case 51%). This comparison between studies is hampered because of differences in the definition of the reference group.

Up to now, studies have provided limited information on the validity of health claims data in specific subgroups. Only one of the four previous studies included patients <65 years of age [4].

That study showed a higher sensitivity in patients with advanced CKD under the age of 65 years compared with patients >65 years (sensitivity 85.8% versus 68.1%). Our data show a similar trend, with a sensitivity considerably higher in patients <75 years compared with patients ≥75 years.

It is not surprising that health claims data in the Netherlands have low accuracy for the estimation of the CKD prevalence in the general population and in particular for elderly patients. In the Netherlands, only hospital claims include information on diagnosis while primary care claims do not. As a consequence, one can only detect CKD patients referred to a nephrologist, and not CKD patients treated by the GP. Patients with advanced CKD have an indication for referral, while the majority of CKD patients in earlier stages are cared for in primary care, especially at older age [12, 13]. This also holds true for many end-stage kidney disease (ESKD) patients on comprehensive conservative management. The results of our study indeed indicate that in daily practice elderly patients with an impaired kidney function are more often treated by a GP or do not receive specific nephrology-related care at all [14]. Of note, with this health claims database, we can demonstrate that adequate nephrological care is registered for 91% of advanced CKD patients aged <75 years.

Our study shows that considerably fewer elderly women with advanced CKD could be identified with health claims data than similarly aged men. It is known that sex differences exist in the epidemiology and outcomes of CKD. Studies show that more women than men have CKD (not on renal replacement therapy) while men show a faster decline in kidney function and more often progress to ESKD [15]. Although current guidelines do not involve sex-specific recommendations in the treatment of CKD, this study suggests that at least in our study sample elderly women with advanced CKD were less likely to be

treated by a nephrologist than men, possibly because elderly women are more likely to choose comprehensive conservative management, which can also be done by a GP, than men [16, 17].

Overall, sensitivity differs considerably across patient subgroups defined by severity of kidney disease, age and sex. This could possibly suggest that clinicians, among other things, take into account an individual's lifetime risk of developing ESKD while considering the need for nephrological care [18]. This risk estimation is among other things based on a person's age, sex and the severity of renal failure. As a result, particularly young patients, men and advanced CKD (Stages 4–5) patients satisfying the chronicity criterion are known within the confines of nephrological care and can thus be identified using health claims data.

Estimating CKD prevalence with populations surveys versus health claims data

Numerous studies have evaluated the prevalence of CKD using population surveys [19, 20], showing that CKD prevalence varies widely with estimations of CKD Stages 3–5 prevalence in Europe varying between 1.0% and 5.9% [20], and in the Netherlands ranging from 1.3% to 4.8% [21, 22]. However, an accurate comparison of CKD prevalence across studies remains challenging since different studies used different CKD definitions and different methods for the assessment of kidney function [23, 24]. Moreover, these studies are always based on samples from the general population. Therefore, when estimating CKD prevalence in population surveys, sampling bias cannot be avoided.

The unadjusted (eGFR-based) CKD prevalence (eGFR <60 mL/min/1.73 m^2) in previous studies using health claims data ranged from 19% in a sample of the general population [4] to 67% using patients hospitalized for myocardial infarction [6]. Since studies use different methods as the reference group, comparison between studies is difficult. The estimated unadjusted prevalence of CKD Stages 3–5 of 22% in our study, using a regional laboratory for the CKD diagnosis, approximates the prevalence of other studies using a sample of the general population as a reference.

Our results suggest that health claims data have low sensitivity for the estimation of overall CKD in the general population, especially in the case of elderly CKD patients and patients with less advanced CKD. However, our results also indicate that health claims data may have value in estimating CKD prevalence in specific subgroups, particularly in young patients and those with advanced CKD. In addition, the sensitivity of young patients (20–59 years) with advanced CKD is similar to those described in validity studies testing claims data for the identification of dialysis patients [25–27], a population for which is generally assumed that health claims data provide reliable estimates of the actual population receiving dialysis treatment.

Strengths and limitations

The strength of our study is the availability of a large laboratory database including all adults with a serum creatinine measurement, allowing its use as the reference group. This enabled us to define CKD in two ways: based on a single and on ≥ 2 creatinine measurements in accordance with the chronicity criterion. We consider ≥ 2 measurements as optimal since it is in accordance with the clinical guidelines. In addition, we were able to differentiate between patients with an eGFR <60 mL/min/1.73 m^2 and an eGFR <30 mL/min/1.73 m^2, and by age and sex.

Several limitations of our study also need consideration. First, primary care claims, in contrast to hospital claims, do not include diagnosis information. Therefore, CKD patients treated by a GP cannot be detected through reimbursement data. Moreover, in the Netherlands, a referral from the GP is always required to consult a medical specialist and therefore Dutch health claims data represent those patients with an indication for referral. Although the outpatient population was considered the best proxy for the general population, the CKD prevalence of individuals treated by the GP and with undetected CKD remains unknown. In this study, we focus on CKD Stages 3–5, since there is no specific health claim for earlier CKD stages and these patients are often undetected or are cared for in primary care. Secondly, the unadjusted CKD prevalence in our database estimated with eGFR was 21.5% for CKD Stages 3–5 and 3.4% for CKD Stages 4–5 and decreased to 13.0% and 2.2%, respectively, after adjustment for age and sex. This means that in our study population elderly individuals were over-represented. This can be expected since we select persons with a performed laboratory test, who are likely to be older than persons from the general population. The unadjusted CKD prevalence estimated by claims data was 6.1% for CKD Stages 3–5 and 2.2% for CKD Stages 4–5 and decreased to 4.0% and 1.5% after adjustment for age and sex. Likely the CKD prevalence estimated with claims data is underestimated as this study shows that the overall sensitivity is low. Finally, the results of this hospital-based study may not be generalizable to a national level due to differences in coding between regions or hospitals in the Netherlands. In addition, generalizability of the results to other countries could be hampered by differences in coding for claims in different healthcare systems.

CONCLUSION

This study shows that the sensitivity of the claim-based diagnoses of CKD and advanced CKD varies largely across patient subgroups. Although overall sensitivity was low, in general, sensitivity was much higher in young patients compared with elderly patients and higher in men than in women. Moreover, health claims data were more accurate in the identification of patients with advanced CKD than those with CKD.

When using health claims data for the estimation of CKD prevalence, it is important to take into account the characteristics of the population at hand. According to this study, the younger the subjects and the more advanced the stage of CKD the higher the sensitivity of such data. Understanding which patients are selected using health claims data and which patients are not is crucial for a correct interpretation of study results.

Bearing this in mind and considering their specific advantages health claims data can have added value for the monitoring of trends in disease prevalence and healthcare costs over time. The linkage of health claims databases to other administrative databases or clinical data can result in a more accurate identification of CKD patients and could thereby improve the usage and value of health claims data for health research even more [28].

SUPPLEMENTARY DATA

Supplementary data are available at ckj online.

REFERENCES

1. Mues KE, Liede A, Liu J et al. Use of the Medicare database in epidemiologic and health services research: a valuable source of real-world evidence on the older and disabled populations in the US. Clin Epidemiol 2017; 9: 267–277
2. Motheral BR, Fairman KA. The use of claims databases for outcomes research: rationale, challenges, and strategies. Clin Ther 1997; 19: 346–366
3. Birnbaum HG, Cremieux PY, Greenberg PE et al. Using healthcare claims data for outcomes research and pharmacoeconomic analyses. PharmacoEconomics 1999; 16: 1–8
4. Ronksley PE, Tonelli M, Quan H et al. For the Alberta Kidney Disease Network. Validating a case definition for chronic kidney disease using administrative data. Nephrol Dial Transplant 2012; 27: 1826–1831.
5. 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
6. Winkelmayer WC, Schneeweiss S, Mogun H et al. Identification of individuals with CKD from Medicare claims data: a validation study. Am J Kidney Dis 2005; 46: 225–232
7. Muntner P, Gutierrez 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: 249–258
8. 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
9. 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
10. Mohnen SM, van Oosten MJM, Los J et al. Healthcare costs of patients on different renal replacement modalities – analysis of Dutch health insurance claims data. PLoS One 2019; 14: e0220800
11. Westerdijk M, Zuurbier J, Ludwig M et al. Defining care products to finance health care in the Netherlands. Eur J Health Econ 2012; 13: 203–221
12. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney International Supplement 2013; 3: 1–150.
13. Vassalotti JA, Centor R, Turner BJ et al. Practical approach to detection and management of chronic kidney disease for the primary care clinician. Am J Med 2016; 129: 153–162.e7
14. Van Gelder VA, Scherpbier-De Haan ND, De Grauw WJC et al. Quality of chronic kidney disease management in primary care: a retrospective study. Scand J Prim Health Care 2016; 34: 73–80
15. Carrero JJ, Hecking M, Chesnaye NC et al. Sex and gender disparities in the epidemiology and outcomes of chronic kidney disease. Nat Rev Nephrol 2018; 14: 151–164
16. Chandna SM, Carpenter L, Da Silva-Gane M et al. Rate of decline of kidney function, modality choice, and survival in elderly patients with advanced kidney disease. Nephron 2016; 134: 64–72
17. Morton RL, Turner RM, Howard K et al. Patients who plan for conservative care rather than dialysis: a national observational study in Australia. Am J Kidney Dis 2012; 59: 419–427
18. Turin TC, Tonelli M, Manns BJ et al. Lifetime risk of ESRD. J Am Soc Nephrol 2012; 23: 1569–1578
19. Hill NR, Fatoba ST, Oke JL et al. Global prevalence of chronic kidney disease – a systematic review and meta-analysis. PLoS One 2016; 11: e0158765
20. Bruck K, Stel VS, Gambaro G et al. CKD prevalence varies across the uropean general population. J Am Soc Nephrol 2016; 27: 2135–2147
21. de Zeeuw D, Hillege HL, de Jong PE. The kidney, a cardiovascular risk marker, and a new target for therapy. Kidney International Supplement 2005; 68: S25–S29
22. van der Ende MY, Hartman MHT, Hagemeijer Y et al. The LifeLines Cohort Study: prevalence and treatment of cardiovascular disease and risk factors. Int J Cardiol 2017; 228: 495–500
23. Bruck K, Jager KJ, Dounousi E et al. Methodology used in studies reporting chronic kidney disease prevalence: a systematic literature review. Nephrol Dial Transplant 2015; 30 (Suppl 4): iv6–iv16
24. Stanifer JW, Muiru A, Jafar TH et al. Chronic kidney disease in low- and middle-income countries. Nephrol Dial Transplant 2016; 31: 868–874
25. Clement FM, James MT, Chin R et al. the Alberta Kidney Disease Network. Validation of a case definition to define chronic dialysis using outpatient administrative data. BMC Med Res Methodol 2011; 11: 25.
26. Komenda P, Yu N, Leung S et al. Determination of the optimal case definition for the diagnosis of end-stage renal disease from administrative claims data in Manitoba, Canada. CMAJ Open 2015; 3: E264–E269
27. Taneja C, Berger A, Inglese GW et al. Can dialysis patients be accurately identified using healthcare claims data? Perit Dial Int 2014; 34: 643–651
28. Slobbe LC. Working with Administrative Health Data - Finding Solid Ground in the Data Morass, Doctoral Thesis, Ipskamp, Tilburg University, 2019. 203 p. https://research.tilburguniversity.edu/en/publications/working-with-administrative-health-data-finding-solid-ground-in-t