Health claims databases used for kidney research around the world

van Oosten, Manon J. M.; Logtenberg, Susan J. J.; Edens, Mireille A.; Hemmelder, Marc H.; Jager, Kitty J.; Bilo, Henk J. G.; Stel, Vianda S.

Published in:
Clinical Kidney Journal

DOI:
10.1093/ckj/sfaa076

Publication date:
2021

Clinical Kidney Journal, 2021, vol. 14, no. 1, 84–97

doi: 10.1093/ckj/sfaa076
Advance Access Publication Date: 27 June 2020
CKJ Review

CKJ REVIEW

Health claims databases used for kidney research around the world

Manon J.M. van Oosten[1], Susan J.J. Logtenberg[2], Mireille A. Edens[3], Marc H. Hemmelder[4,5], Kitty J. Jager[1], Henk J.G. Bilo[3,6,7] and Vianda S. Stel[1]

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

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

ABSTRACT

Health claims databases offer opportunities for studies on large populations of patients with kidney disease and health outcomes in a non-experimental setting. Among others, their unique features enable studies on healthcare costs or on longitudinal, epidemiological data with nationwide coverage. However, health claims databases also have several limitations. Because clinical data and information on renal function are often lacking, the identification of patients with kidney disease depends on the actual presence of diagnosis codes only. Investigating the validity of these data is therefore crucial to assess whether outcomes derived from health claims data are truly meaningful. Also, one should take into account the coverage and content of a health claims database, especially when making international comparisons. In this article, an overview is provided of international health claims databases and their main publications in the area of nephrology. The structure and contents of the Dutch health claims database will be described, as well as an initiative to use the outcomes for research and the development of the Dutch Kidney Atlas. Finally, we will discuss to what extent one might be able to identify patients with kidney disease using health claims databases, as well as their strengths and limitations.

Keywords: CKD, dialysis, epidemiology, health claims data, health claims database, kidney transplantation

INTRODUCTION

There are many registries and studies collecting information on patients with chronic kidney disease (CKD) and patients on renal replacement therapy (RRT). In recent decades, these have been supplemented by administrative healthcare data, including health insurance claims data, thereby providing new research opportunities.

Health claims data are routinely collected for payment purposes, and for this purpose they usually are comprehensive and complete. These generally contain sociodemographic data and

Received: 11.2.2020; Editorial decision: 7.4.2020

longitudinal data on medical diagnoses and procedures, pharmacological treatment and costs. Most of these databases have nationwide coverage, include all age categories and offer data that reflect day-to-day clinical practice. However, when using administrative data for healthcare research, it is important to recognize their unique opportunities as well as their limitations. Furthermore, for international comparisons, one should take into consideration the differences in database characteristics when analysing and interpreting these data [1]. Nowadays, several health claims databases of national and regional healthcare systems are available and are being used for kidney research.

In this article we provide a worldwide overview of health claims databases and summarize their main publications in the area of nephrology. Next, we will introduce the Dutch health claims database and a new initiative to use it for research and the development of a Dutch Kidney Atlas. Finally, we will discuss to what extent they allow identification of patients with kidney disease as well as the strengths and limitations of such databases.

WORLDWIDE HEALTH INSURANCE CLAIMS DATABASES IN KIDNEY RESEARCH

We selected papers in which health claims data, whether or not combined with other administrative databases, were used for research on patients with kidney disease. A systematic literature search was not possible because the use of health claims data is often not clearly mentioned in the articles. We performed an Internet and PubMed search in August 2019 using different search terms (e.g. health claims data, health insurance claims data, healthcare claims data, administrative data in combination with terms to search for research on patients with kidney disease such as CKD, dialysis or kidney transplantation). We selected suitable papers based on the abstract and methods section. In addition, we checked the references of selected papers and checked other published papers by the main authors.

We identified 13 health claims databases in 10 countries (Canada, China, the UK, France, India, Japan, South Korea, Taiwan, The Netherlands and the USA) that were used for research on patients with kidney disease and resulted in at least one publication in English in a scientific journal (Table 1). Other countries, such as Austria and Germany, have a health claims database that is used for clinical research. However, we were unable to find papers published in scientific journals specifically focusing on kidney disease patients from these countries [2]. In some countries in Southeast Asia, health claims databases have been recently established and so far their use for research purposes has been rare [3]. The use of health claims databases for kidney research worldwide may therefore be expanded.

The organization of healthcare systems, healthcare financing and socio-economic settings differ importantly between countries [4]. This results in differences in the coverage, the size and the content of the health claims databases. A recent study presented a comprehensive overview of existing systems and financing for kidney care and demonstrated significant heterogeneity [5]. In countries where health insurance is obligatory or universally accessible (suh as Canada, France, Japan, South Korea, The Netherlands and Taiwan), health claims databases have (almost) complete coverage of all country or province inhabitants (Table 1). The coverage of the Chinese claims databases, in contrast, is much less complete. Although they comprise a large number of individuals, the content of these is limited to the ones with access to healthcare insurance. The USA does not have universal healthcare coverage and the Medicare program only provides health insurance for American citizens ≥65 years of age as well as for people of all ages with severe diseases such as those with end-stage kidney disease. Therefore, the Medicare claims database covers a selection of the US population. Furthermore, the Rajiv Aarogyasri Community Health Insurance Scheme is an Indian state government program providing free hospital care to poor individuals. Noteworthy is the Hospital Episode Statistics (HES) database, which provides admission data for National Health Service (NHS) hospitals in the UK.

Within health claims data, kidney disease patients are generally identified with International Classification of Disease codes in combination with medical procedure codes. In Japan, patients are identified with Diagnosis Procedure Codes (DPC), whereas in The Netherlands patients are identified using diagnosis treatment combinations, a system similar to the DPC system.

When evaluating studies using health claims data, it is essential to consider the coverage (such as age distribution, health system characteristics, insurance coverage and percentage of the population with healthcare insurance) and the specific procedure of patient selection to correctly interpret its results. This is of special importance when making international comparisons.

Accessibility of health claims databases

Several barriers for the secondary use of administrative data in general have been identified. The impact of these barriers differs considerably between countries and influences the availability and utility of health claims data for research per country [6]. In general, the complexity and the enormous amount of unprocessed raw data make it difficult and laborious to work with health claims databases. The pre-processing of data is time-consuming and requires experience with working with big data. In addition, extensive application processes as well as data processing fees are usually needed to access the data.

Data accessibility differs across countries. For example, the National Health Insurance Research Database (NHIRD) of Taiwan is known for its high accessibility. The database is publicly available and any researcher may apply for data that are provided for a small processing fee [7]. In other countries (e.g. South Korea, Japan, France, The Netherlands, Canada, the UK and the USA), researchers have to go through more extensive data request procedures [8–14]. Physical access to extracts of de-identified data is sometimes only possible in a designated secure environment (The Netherlands) or researchers are required to have a secured physical environment at their institution (Japan) [9, 15]. In France, data are supplied on a secure, electronic medium while in Canada access is provided via a secure, online research environment [10, 12]. Information about the accessibility of the databases of China and India was not available on public websites or in publications [8].

The processing fees are determined by fees for data extraction, the complexity of the data request and possible extra delivered services such as professional assistance in working with the complex databases. The level of the fees is therefore difficult to determine, but they are often described as costly. The extra financing needed for the use of health claims databases as well as the expertise needed for a study on health claims data (e.g. data analyst specialized in big data, nephrologist, epidemiologist and PhD student) may be challenging, especially for

Table 1. Overview of health claims databases in the world used for kidney research

Country or province	Number of inhabitants in 2018	Health claims database	Coverage	Linkage to other (administrative) databases
Canada - Alberta	4.3 million	Alberta Provincial Physician Claims database	>99% of inhabitants	Part of the Alberta Kidney Disease Network database; linkage to the Northern and Southern Alberta Renal Programs and clinical laboratory data
Canada - Manitoba	1.4 million	Manitoba Health Physician Claims database	>99% of inhabitants	Linkage to Manitoba Renal Program Dialysis Registry
Canada - Ontario	14.3 million	Ontario Health Insurance Plan database (OHIP)	>99% of inhabitants	Linkage to Ontario's central organ and tissue donation agency, Canadian Institute for Health Information Discharge Abstract Database (CIHI-DAD), National Ambulatory Care Reporting System (data on emergency room visits), Ontario Registered Persons Database Information (demographics and vital status), Ontario Drug Benefit Plan (outpatient prescription drug usage for individuals ≥65 years)
Canada - Quebec	8.4 million	Régie de l'assurance maladie du Québec (RAMQ)	>99% of inhabitants	Linkage to Canadian Organ Replacement Register (CORR)
China	1.4 billion	China Health Insurance Research Association database (CHIRA)	977 million insured people in 2015	
		Commercial Health Insurance database (CHI)	60 million customers in 2015	
UK	55.3 million	Hospital Episode Statistics (HES)	All admissions to NHS hospitals in the UK	Linkage to the Office for National Statistics (ONS) for mortality data
France	66.9 million	Système national d'information interrégimes de l'Assurance Maladie (Sniiram)	96% of inhabitants	Linkage to French Renal Epidemiology and Information Network (REIN) registry, French national hospital computerized medical information system (PMSI)
India - Andhra Pradesh	1.3 billion	Rajiv Aarogyasri Community Health Insurance Scheme	81% of inhabitants	
Japan	127.2 million	National Database of Health Insurance Claims and Specific Health Checkups of Japan (NDB)	90% of inhabitants	
South Korea	51.1 million	Health Insurance and Review Assessment Service (HIRA)	98% of inhabitants	Linkage to a national health screening program (including 10 million Koreans) providing information on serum creatinine and urine albumin measurements
Taiwan	23.8 million	National Health Insurance Administration Research Database (NHIRD)	>99% of inhabitants	Linkage to e.g. death registry, cancer registry, reportable infectious disease registry
The Netherlands	17.1 million	Vektis database	98% of inhabitants	
USA	327.2 million	Medicare Services	All patients on RRT aged ≥65 years and patients with end-stage renal disease (RRT)	Linkage to the United States Renal Data System (USRDS), Scientific registry of transplant recipients (SRTS), National Health and Nutrition Examination Survey data (NHANES)

grant applications that use a fixed amount and usually mainly reimburse the salary costs of the researcher.

Studies using health claims data

Health claims data are being used to answer a wide variety of healthcare-related questions in different fields of research. Several countries have the ability to link their health claims database to other administrative databases such as national dialysis and transplantation registries, national vital statistics or clinical laboratory data (Table 1). This linkage capacity greatly expands its research possibilities [16].

In general, studies using health claims data mainly focus on the epidemiology of a disease, including patients' morbidity and

survival, evaluating healthcare costs and the delivery of healthcare services, describing prescription patterns and the effectiveness of pharmacological therapies and exploring clinical outcomes [14]. Table 2 provides a selection of papers on patients with kidney disease based on health claims data (or a combination of health claims data and other linked administrative data) published in scientific journals. We divided these articles into three main fields of research, i.e. validation studies (see Table 3), cost studies and descriptive and outcome studies, to provide insights into the variety of study questions.

Most papers reported on patients receiving dialysis treatment and significantly less on patients with kidney transplantation or CKD patients not treated with RRT. Interestingly, health claims data can also provide opportunities to report on specific kidney diseases such as polycystic kidney disease [54]. Several papers focused on the validity of health claims data to identify patients with CKD [57–59], dialysis [60–62] and kidney transplantation [63]. The results of these validity studies are presented in Table 3 and will be discussed later.

As health claims databases include reimbursement data, it is possible to study healthcare costs. Several studies provided cost estimations for dialysis and kidney transplantation in different countries [17–21, 23–26]. Estimations of healthcare costs of CKD patients not treated with RRT were less common and only provided in two studies [15, 22]. Since health claims usually contain longitudinal data with nationwide coverage, studies frequently report epidemiological data such as incidence and prevalence of kidney diseases, the occurrence of kidney disease–related risk factors/comorbidities, mortality rates and their trends over time. A wide range of outcomes have been described using health claims data. Frequently studied outcomes include those that are cardiovascular related, such as cardiovascular disease in transplant patients, stroke, atrial fibrillation and major cardiovascular events. The comprehensive drug delivery data of pharmacies provide unique opportunities for pharmaco-epidemiological research to study the use and effects of medication in CKD patients. This results, for example, in studies reporting the use of warfarin, angiotensin-converting enzyme inhibitors/angiotensin receptor blockers or β-blockers in dialysis patients or metformin use in CKD patients. With pharmacological data, it is important to realize that non-reimbursable drugs or over-the-counter drugs remain undetected in health claims data. Other less common outcomes reported include depressive disorder in kidney transplantation, fracture risk and associated mortality in kidney transplantation, the peptic ulcer rebleeding risk in dialysis patients or the risk of end-stage renal disease after hypertensive disorders in pregnancy (see Table 2).

The number of published papers on kidney disease patients differs widely by database, with Taiwan being the leading country in utilizing claims data for clinical research. Data from Taiwan's NHIRD is available for any researcher in Taiwan for a small processing fee, which has been shown to increase the publication rate dramatically. This emphasizes the importance of keeping the financial and technical barriers for the reuse of health claims data for research purposes as low as possible [64].

The Dutch Kidney Atlas

Recently, Dutch health claims data have been processed to study patients with kidney disease in the Dutch Kidney Atlas project. Box 1 provides a detailed description of the Dutch health claims database (called Vektis) and the related healthcare system. The Dutch Kidney Atlas project provides information on the number of CKD patients {CKD Stage G4–G5 [estimated glomerular filtration rate (eGFR) <30 mL/min/1.73 m^2]} not treated with RRT, those on dialysis and kidney transplant patients, their healthcare costs, prescribed medication, outcomes (such as the number of hospital visits, intensive care unit admittance and mortality) and comorbid disorders like diabetes mellitus, hypertension and cardiovascular diseases. Data are published on a website (www.nieratlas.nl) and are reported by age group and sex and are compared with a reference group from the general population. Figure 1 shows several graphs presented in the Dutch Kidney Atlas. Data are presented on a national level as well as on a regional level to demonstrate potential geographic variation. Furthermore, the website includes data since 2012 and will be updated on an annual basis. The website was designed for public use by healthcare professionals, policymakers, researchers and insurance companies, as well as patients with CKD.

The Dutch Kidney Atlas project also involves scientific research on patients with CKD using the Dutch health claims database. So far, two studies on the healthcare costs of patients with CKD with and without RRT have been published [15, 25].

Box 1. Dutch health claims database

- In The Netherlands, healthcare provision and payment for healthcare and healthcare-related services through insurance is embedded within a social security system [65]. Because basic health insurance is obligatory for all Dutch residents, an estimated 99.8% [66] of the Dutch population of ∼17 million people has healthcare insurance [67]. The Dutch healthcare system has a gatekeeping principle, which means that patients can easily contact a primary care provider (e.g. general practitioner, dentist, midwife and physiotherapist), but hospital care and specialist care require a referral from a primary care provider, with emergency care as an exception.
- Basic health insurance covers the main aspects of healthcare, including primary care, hospital care, medication, mental healthcare, maternity care and home nursing care. Care not covered by the basic insurance can be insured through voluntary health insurance. Health insurance companies pay the hospital based on Diagnose Behandelcombinatie (DBCs), a system similar to the concept of diagnosis-related groups. A DBC contains information characterizing the delivered hospital care for a specific medical condition or complaint by type of specialization. The DBC comprises all medical activities needed, from establishing the diagnosis to the last check after treatment, and thereby describing a complete care episode. Every type of DBC has a fixed price, which is the sum of costs of all intermediate products, i.e. the activities, thereby covering all direct and indirect costs of a care episode [68].
- The health claims data of all Dutch health insurance companies are collected in the Vektis database, which covers (almost) all inhabitants of The Netherlands. For each health claim in the Vektis database, data are available on patient characteristics (year of birth, sex, area of residence, socio-economic status and date of death) and the costs involved [16]. Vektis complies with the Dutch law and the European General Data Protection Regulation. To ensure privacy while performing the present research, Vektis pseudonymized the persons' national identification number and data access is only allowed in a physically secured environment designated by Vektis; only aggregated data are allowed to leave this secured environment. All contributing insurance companies provided permission for the use of these national data.

Table 2. Selection of papers on kidney disease patients based on health claims data published in scientific journals

Study types	Author	Journal	Year	Country/region	Population	Title
Cost studies	Chang *et al.* [17]	*Nephrology*	2015	Taiwan	Dialysis	Trends of cost and mortality of patients on haemodialysis with end stage renal disease
	Chang *et al.* [18]	*Sci Rep*	2016	Taiwan	Dialysis	Cost-effectiveness of haemodialysis and peritoneal dialysis: a national cohort study with 14 years follow-up and matched for comorbidities and propensity score
	Couchoud *et al.* [19]	*Nephrol Dial Transplant*	2015	France	Dialysis, transplantation	Economic impact of a modification of the treatment trajectories of patients with end-stage renal disease
	Couillerot-Peyrondet *et al.* [20]	*Eur J Health Econ*	2017	France	Dialysis, transplantation	A comprehensive approach to assess the costs of renal replacement therapy for end-stage renal disease in France: the importance of age, diabetes status, and clinical events
	Helmuth *et al.* [21]	*Clin J Am Soc Nephrol*	2019	USA	Transplantation	Secular trends in the cost of immunosuppressants after solid organ transplantation in the United States
	Honeycutt *et al.* [22]	*J Am Soc Nephrol*	2013	USA	CKD	Medical costs of CKD in the Medicare population
	Kao *et al.* [23]	*Perit Dial Int*	2013	Taiwan	Dialysis	Lifetime costs for peritoneal dialysis and haemodialysis in patients in Taiwan
	Kitazawa *et al.* [24]	*Transplant Proc*	2017	Japan	Transplantation	Cost analysis of transplantation in Japan, performed with the use of the National Database
	Mohnen *et al.* [25]	*PLoS One*	2019	Netherlands	Dialysis, transplantation	Healthcare costs of patients on different renal replacement modalities – analysis of Dutch health insurance claims data
	van Oosten *et al.* [15]	*Nephrol Dial Transplant*	2019	Netherlands	CKD, dialysis, transplantation	Age-related difference in healthcare use and costs of patients with chronic kidney disease and matched controls: analysis of Dutch health claims data
	Shaikh *et al.* [26]	*Kidney Int*	2018	India	Dialysis	Utilization, costs, and outcomes for patients receiving publicly funded haemodialysis in India
Descriptive and outcome studies	Chettiar *et al.* [27]	*Clin J Am Soc Nephrol*	2018	USA	Dialysis, transplantation	Association of inpatient palliative care with health care utilization and postdischarge outcomes among Medicare beneficiaries with end stage kidney disease
	Choi *et al.* [28]	*Am J Nephrol*	2017	South Korea	Dialysis, transplantation	Disparities in kidney transplantation access among Korean patients initiating dialysis: a population-based cohort study using national health insurance data (2003-2013)
	Dobbels *et al.* [29]	*Am J Kidney Dis*	2008	USA	Transplantation	Depressive disorder in renal transplantation: an analysis of Medicare claims
	Farrugia *et al.* [30]	*Kidney Int*	2014	UK	Transplantation	Malignancy-related mortality following kidney transplantation is common
	Ferreira *et al.* [31]	*Nephrol Dial Transplant*	2019	France	Dialysis	Angiotensin-converting enzyme inhibitors/angiotensin receptor blockers, β-blockers or both in incident end-stage renal disease patients without cardiovascular disease: a propensity-matched longitudinal cohort study.
	Ferro *et al.* [32]	*Clinical Transplant*	2015	UK	Transplantation	Fracture risk and mortality post-kidney transplantation
	Han *et al.* [33]	*Clin J Am Soc Nephrol*	2015	South Korea	Dialysis	Dialysis modality and mortality in the elderly: a meta-analysis
	Hayer *et al.* [34]	*Diabetologia*	2014	UK	Transplantation	Infection-related mortality is higher for kidney allograft recipients with pretransplant diabetes mellitus
	Hung *et al.* [35]	*Lancet Diabetes Endocrinol*	2015	Taiwan	CKD	Metformin use and mortality in patients with advanced chronic kidney disease: national, retrospective, observational, cohort study

(continued)

Table 2. Continued

Study types	Author	Journal	Year	Country/region	Population	Title
	Kim *et al.* [36]	*Int J Cardiol*	2015	South Korea	Dialysis	Risk of major cardiovascular events among incident dialysis patients: a Korean national population-based study
	Kitchlu *et al.* [37]	*Nephrol Dial Transplant*	2012	Canada, Ontario	Dialysis	Beta-blockers and cardiovascular outcomes in dialysis patients: a cohort study in Ontario, Canada
	Komenda *et al.* [38]	*CMAJ Open*	2015	Canada, Manitoba	Dialysis	Secular trends in end-stage renal disease requiring dialysis in Manitoba, Canada: a population-based study
	Kuo *et al.* [39]	*Am J Kidney Dis*	2007	Taiwan	CKD	Epidemiological features of CKD in Taiwan
	Lam *et al.* [40]	*Transplantation*	2017	Canada, Ontario	Transplantation	The risk of cardiovascular disease is not increasing over time despite aging and higher comorbidity burden of kidney transplant recipients
	Lenihan *et al.* [41]	*Transplantation*	2019	USA	Transplantation	Trends in the medical complexity and outcomes of Medicare-insured patients undergoing kidney transplant in the years 1998–2014
	Li *et al.* [42]	*Nephrol Dial Transplant*	2012	Taiwan	Transplantation	Malignancies after renal transplantation in Taiwan: a nation-wide population-based study
	Liao *et al.* [43]	*Kidney Int*	2015	Taiwan	Dialysis	Incidence and risk factors for new-onset atrial fibrillation among patients with end-stage renal disease undergoing re-nal replacement therapy
	René *et al.* [44]	*Nephrol Dial Transplant*	2017	Canada, Quebec	Dialysis	Association of erythropoiesis-stimulating agents and the inci-dence risk of cancer diagnosis among chronic dialysis patients: a nested case–control study
	Tonelli *et al.* [45]	*Kidney Int*	2018	Canada, Alberta	CKD	A population based cohort study defines prognoses in severe chronic kidney disease
	Wang *et al.* [46]	*Kidney Int*	2019	China	CKD, dialysis, transplantation	Executive summary for the 2015 Annual Data Report of the China Kidney Disease Network (CK-NET)
	Wang *et al.* [47]	*Am J Kidney Dis*	2014	Taiwan	Dialysis	Risk of stroke in long-term dialysis patients compared with the general population
	Wang *et al.* [48]	*Clin J Am Soc Nephrol*	2015	Taiwan	Dialysis	Comparison of subdural haematoma risk between haemodialy-sis and peritoneal dialysis patients with ESRD
	Wang *et al.* [49]	*Nephrol Dial Transplant*	2018	Taiwan	Dialysis	Risk of new-onset diabetes in end-stage renal disease patients undergoing dialysis: analysis from registry data of Taiwan
	Weinhandl *et al.* [50]	*Am J Nephrol*	2019	USA	Dialysis	Contemporary trends in clinical outcomes among dialysis patients with Medicare coverage
	Wu *et al.* [51]	*Gut*	2011	Taiwan	Dialysis	Long-term peptic ulcer rebleeding risk estimation in patients undergoing haemodialysis: a 10-year nationwide cohort study
	Wu *et al.* [52]	*Am J Obstet Gynecol*	2014	Taiwan	Dialysis	End-stage renal disease after hypertensive disorders in pregnancy
	Yoon *et al.* [53]	*Stroke*	2017	South Korea	Dialysis	Warfarin use in patients with atrial fibrillation undergoing haemodialysis
	Yu *et al.* [54]	*Lancet Oncol*	2016	Taiwan	Polycystic kidney disease	Risk of cancer in patients with polycystic kidney disease: a pro-pensity-score matched analysis of a nationwide, population-based cohort study
Validation studies	For more details see Table 3					

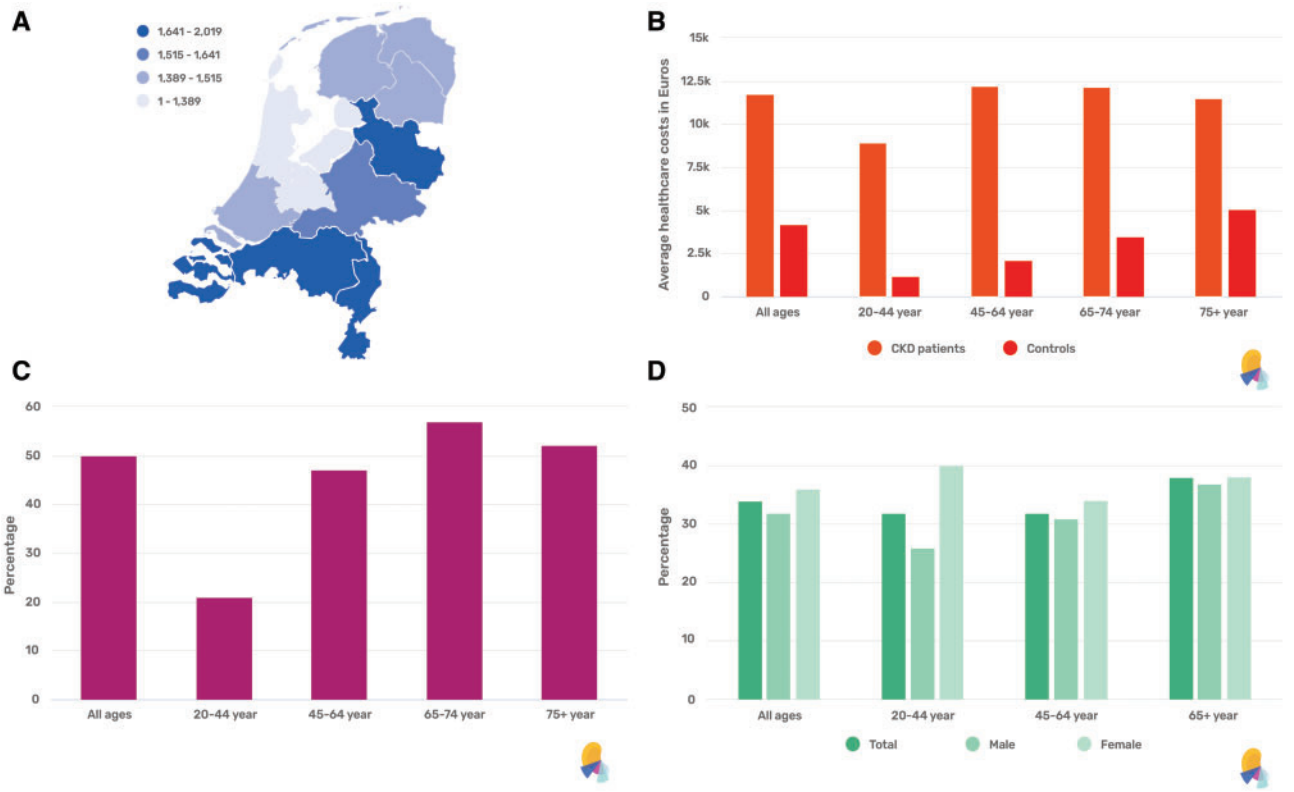

FIGURE 1: Examples of the Dutch Kidney Atlas. (**A**) Geographical variation of the number of patients with CKD Stages G4–G5 (diagnosis code eGFR <30 mL/min/1.73 m^2) not treated with RRT by province of The Netherlands, 2017; numbers per million insured population. (**B**) Total healthcare costs (€) of patients with CKD Stages G4–G5 (diagnosis code eGFR <30 mL/min/1.73 m^2) not treated with RRT compared with a matched control group (2017, presented for the total group and different age groups). (**C**) Statin use in prevalent dialysis patients (2017, percentage of the total population), presented for the total group and different age groups. (**D**) Percentage of kidney transplant patients visiting the emergency department per year (2017), presented for the total group and subgroups based on age and gender.

IDENTIFICATION OF PERSONS WITH KIDNEY DISEASE

Administrative databases are the result of administrative processes, such as reimbursement in the case of health claims databases, and are therefore not designed for clinical research purposes [16]. Because clinical data and information on renal function are lacking in health claims databases, the identification of patients with kidney disease depends on diagnosis codes only, which in turn depends on the proper registration of the codes by the involved healthcare professional or organization. Investigating the accuracy of this identification is therefore crucial to assess whether data derived from health claims data are truly representative and this information on the validity of the diagnosis codes should be provided in research articles using health claims data [58, 69]. Some studies assessed the validity of health claims data to identify patients with CKD, dialysis or kidney transplantation (Table 3). Apart from one study using Dutch health claims data, all studies were performed in Canada or the USA. Of all available studies, five studied the validity of health claims data in the identification of CKD patients, three of dialysis patients and one of kidney transplant patients. However, the use of different case definitions and reference populations does impede the direct comparison between studies. In the following paragraphs we will discuss the studies that have assessed the validity of health claims data to identify patients with CKD, dialysis or kidney transplantation in more detail.

CKD

A study in Alberta, Canada, tested health claims data for the identification of CKD (defined as an eGFR <60 mL/min/1.73 m^2) against a gold standard derived from outpatient serum creatinine measurements. In this study, 19% of CKD patients were correctly identified as such with health claims data (sensitivity or true-positive rate) and 60% of the patients with CKD-related claims data did have an eGFR <60 mL/min/1.73 m^2 [positive predictive value (PPV) 60%] [57]. Similar results were found in a study from Ontario, Canada, using patients with a serum creatinine laboratory test following a prescription of medication as the gold standard (sensitivity 18%, PPV 65%) [55]. In both studies, sensitivity was markedly higher for CKD patients defined as an eGFR <30 mL/min/1.73 m^2 (65% and 59%, respectively). Two US studies tested the validity of Medicare data. One study used patients with hospitalization for myocardial infarction as the gold standard (sensitivity 27%, PPV 89%) [58], while the second study used research study measurements as a reference (sensitivity 16%, PPV 76%) [56]. Recently we tested the validity of Dutch health claims data using a laboratory database as the gold standard. Sensitivity was markedly higher in patients with advanced CKD (eGFR <30 mL/min/1.73 m^2) than in patients with CKD (eGFR <60 mL/min/1.73 m^2), being 51% and 27%, respectively [59]. All studies had high specificity for CKD. The negative predictive value (NPV) varied in all studies and with a wide range (36–98%).

Table 3. Overview of studies of the validity of health claims data in the identification of CKD, dialysis and kidney transplant patients

Author	Health claims database	Study Population	Reference population	Age	Case definition	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
CKD									
Fleet *et al.* [55]	Ontario Health Insurance Plan database (OHIP), Ontario (Canada)	Patients with ICD-10 and physician claims diagnostic codes for CKD, between 1 July 2007 and 31 December 2010	Patients with an outpatient prescription medication and a serum creatinine test the year prior to the prescription date from a laboratory in Southwestern Ontario	≥65 years	eGFR <60 mL/min/1.73 m^2	18.0 (17.7–18.4)	98.2 (98.1–98.3)	85.2 (84.5–85.9)	67.7 (67.4–68.0)
					eGFR <45 mL/min/1.73 m^2	32.7 (32.0–33.3)	96.9 (96.7-97.0)	65.4 (64.4-66.3)	88.8 (88.6–89.0)
					eGFR <30 mL/min/1.73 m^2	58.8 (57.4–60.1)	94.6 (94.5–94.7)	32.5 (31.6–33.5)	98.1 (98.0–98.2)
Muntner *et al.* [56]	Medicare claims data, USA	Patients with ICD-9 discharge codes associated with hospitalization or physician evaluation and claims associated with outpatient physician visits for CKD, between January 2003 and October 2007	Participants enrolled in the Reasons for Geographic and Racial Differences in Stroke (REGARDS) study with data available on eGFR and albumin–creatinine ratio	≥65 years	eGFR <60 mL/min/1.73 m^2 or ACR >30 mg/g	15.5 (14.0–17.1)	97.7 (97.2–98.1)	75.6 (71.4–79.5)	71.5 (70.4–72.6)
					eGFR <60 mL/min/1.73 m^2	20.6 (18.5–22.8)	97.1 (96.6–97.5)	63.9 (59.2–68.3)	83.0 (82.1–83.9)
					eGFR <45 mL/min/1.73 m^2	37.1 (32.7–41.6)	95.8 (95.3–96.2)	39.0 (34.5–43.7)	95.4 (94.9–95.9)
					eGFR <30 mL/min/1.73 m^2	56.4 (45.8–66.6)	94.2 (93.6–94.8)	11.8 (8.9–15.1)	99.4 (99.1–99.5)
Ronksley *et al.* [57]	Alberta Provincial Physician Claims database, Alberta (Canada)	Patients with ICD-9 and ICD-10 codes for CKD, between 1 January 2004 and 31 December 2004	Patients with at least two outpatient serum creatinine measurements within a 1-year time period	≥18 years	eGFR <60 mL/min/1.73 m^2				
					One claim or one hospitalization in one year	18.9 (–)	97.29 (–)	60.59 (–)	83.99 (–)
					Two claims or one hospitalization in 1 year	14.29 (–)	98.19 (–)	63.69 (–)	83.39 (–)
					Three claims or one hospitalization in 1 year	11.89 (–)	98.59 (–)	64.09 (–)	82.99 (–)
					eGFR <30 mL/min/1.73 m^2				
					One claim or one hospitalization in 1 year	64.79 (–)	96.59 (–)	24.09 (–)	99.49 (–)
					Two claims or one hospitalization in 1 year	56.59 (–)	97.79 (–)	29.39 (–)	99.29 (–)
					Three claims or one hospitalization in 1 year	49.99 (–)	98.19 (–)	31.49 (–)	99.19 (–)

(continued)

Table 3. Continued

Author	Health claims database	Study Population	Reference population	Age	Case definition	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
Winkelmayer *et al.* [58]	Medicare claims data, USA	Patients with ICD-9 diagnosis codes for CKD, during 1999 and/or 2000	Patients with hospitalization for myocardial infarction with a serum creatinine measurement	≥65 years	eGFR <60 mL/min/1.73 m^2 for 6-months period	20.7 (18.5–22.9)	96.0 (94.4–97.5)	91.6 (88.5–94.8)	36.3 (34.0–38.7)
					eGFR <60 mL/min/1.73 m^2 for 12-month period	26.6 (24.2–28.9)	93.3 (91.3–95.2)	89.3 (86.3–92.4)	37.4 (35.0–39.8)
van Oosten *et al.* [59]	Vektis database, The Netherlands	Patients with DBC codes for CKD, between 1 January 2014 and 31 December 2014	Patients with an outpatient serum creatinine measurement in 2014	≥18 years	One eGFR <60 mL/min/1.73 m^2	20 (19–21)	100 (100–100)	96 (95–97)	84 (83–84)
					At least two eGFR <60 mL/min/1.73 m^2 at least 90 days apart	27 (25–28)	100 (100–100)	90 (88–91)	52 (51–52)
					One eGFR <30 mL/min/1.73 m^2	42 (38–46)	100 (100–100)	83 (79–87)	98 (98–99)
					At least two eGFR <30 mL/min/1.73 m^2 at least 90 days apart	51 (47–56)	100 (99–100)	80 (76–84)	98 (98–99)
Dialysis									
Clement *et al.* [60]	Alberta Provincial Physician Claims database, Alberta (Canada)	Patients with physician claims for outpatient dialysis, between 1 January 2008 and 31 December 2008	ESRD registry [Northern Alberta (NARP) and Southern Alberta (SARP) registries]	≥18 years	1. At least one claim	81.1 (–)	NA	77.7 (–)	NA
					2. At least two claims	78.6 (–)	NA	80.7 (–)	NA
					3. At least two claims at least 90 days apart	63.1 (–)	NA	84.8 (–)	NA
					4. Continuous claims at least 90 days apart with no gap in claims >21 days	58.2 (–)	NA	85.9 (–)	NA
Komenda *et al.* [61]	Manitoba Health Physician Claims database, Manitoba (Canada)	Patients with physician claims for outpatient dialysis, between 1 January 2004 to 31 December 2010	Manitoba Renal Program Dialysis Registry	>18 years	**1-year period (2010)**				
					1. At least one claim	77.0 (74.7–79.2)	93.8 (92.9–94.7)	85.2 (83.2–87.2)	89.8 (88.7–90.9)
					2. Any two claims	74.6 (72.3–76.9)	94.4 (93. 6–95.2)	86.0 (84.0–88.0)	88.9 (87.8–90.0)
					3. Any two claims at least 90 days apart	64.8 (62.2–67.3)	97.1 (96.5–97.7)	91.2 (89.5–93.0)	85.6 (84.4–86.8)
					4. Any two claims at least 90 days apart with no gaps in treatment >21 days	52.7 (50.1–55.4)	97.5 (96.9–98.1)	90.7 (88.7–92.7)	81.7 (80.4–83.0)
					5-year period (2004-2008)				
					1. At least one claim	87.6 (86.3–89.0)	91.3 (90.7–91.9)	74.4 (72.8–76.0)	96.2 (95.8–96.7)
					2. Any two claims	86.0 (84.7–87.4)	93.4 (92.9–93.9)	78.9 (77.4–80.4)	95.9 (95.4–96.3)

(continued)

Table 3. Continued

Author	Health claims database	Study Population	Reference population	Age	Case definition	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
					3. Any two claims at least 90 days apart	72.0 (70.2–73.8)	99.6 (99.5–99.8)	98.3 (97.7–98.9)	92.5 (92.0–93.1)
					4. Any two claims at least 90 days apart with no gaps in treatment >21 days	47.6 (45.6–49.6)	99.8 (99.7–99.9)	98.3 (97.6–99.0)	86.9 (86.2–87.5)
Taneja *et al.* [62]	Health Alliance Plan (HAP) database, USA	Patients with ESRD and dialysis-related billing codes for peritoneal dialysis or haemodialysis, between 1 January 2005 and 31 December 2008.	Patient medical record	18–63 years	Any PD-related claim—in a 30-day window	NA	NA	34.9 (–)	NA
					Any PD-related claim—in a 90-day window	NA	NA	67.4 (–)	NA
					Any PD-related claim—in a 180-day window	NA	NA	67.4 (–)	NA
					Any HD-related claim—in a 30-day window	NA	NA	86.7 (–)	NA
					Any HD-related claim—in a 90-day window	NA	NA	90.8 (–)	NA
					Any HD-related claim—in a 180-day window	NA	NA	93.1 (–)	NA
Kidney transplantation									
Lam *et al.* [63]	Ontario Health Insurance Plan database (OHIP), Ontario (Canada)	Patients with kidney transplantation related claims, between 1 January 2008 and 31 December 2011	Three major transplant centers in Ontario (Toronto General Hospital, University Hospital – London and Ottawa Hospital)	All	A claim for a kidney-only transplant	98 (97–99)	NA	96 (95–97)	NA

CI, confidence interval; ICD, International Classification of Diseases.

Health claims data have low sensitivity for the estimation of the overall CKD prevalence in the general population, since health claims data only detect CKD patients treated in a hospital and registered for the specific Diagnose Behandelcombinaties (DBCs) and not the ones treated by a general practitioner or those who are not treated at all. However, they do, to a large extent, reflect the population of CKD with an actual referral to a nephrologist.

Dialysis

All available studies indicate that identifying patients undergoing dialysis with health claims data is accurate. A study in Ontario, Canada, showed a reasonably good identification of dialysis patients with a sensitivity of 81% and a PPV of 78% when compared to a gold standard of registry data [60]. These results were confirmed in patients in Manitoba, Canada (sensitivity 77%, PPV 85%), also using registry data as the gold standard [61]. A US study accurately identified haemodialysis using health claims data compared with medical records data as the gold standard (PPV 91%), but was less precise for patients treated with peritoneal dialysis (PPV 67%) [62]. In The Netherlands, the number of dialysis (haemodialysis and peritoneal dialysis) patients identified with Dutch health claims data was compared with the number of patients in the Dutch registry for end-stage renal disease [RENINE (Registratie Nierfunctievervanging Nederland)]. Since the analysis was only possible on aggregated data, sensitivity could not be calculated, but the correspondence between the databases was very high (99%) [25].

Kidney transplantation

Identification of performed kidney transplantations was shown to be very accurate using Canadian health claims data compared with data from three Canadian transplantation centres serving as the gold standard (sensitivity 98%, PPV 96%) [63]. Also, Dutch health claims data were shown to very accurately identify the number of performed kidney transplantations per year when compared with the Dutch registry for end-stage renal disease, with a correspondence of 99% between the databases [25].

In RRT patients, who are generally all treated within the standard healthcare system, both the PPV and NPV are high. This makes it possible to compare outcomes between patients with and without RRT. In contrast, not all affected CKD patients (not treated with RRT) are known or treated within the (hospital) care system. Therefore, in CKD patients, the PPV is high, but the NPV generally is low. In CKD patients, it may therefore be more difficult to identify those without CKD using health claims data [58]. Please note that both the PPV and NPV depend on the prevalence of the disease in the population [70].

OPPORTUNITIES AND CHALLENGES OF HEALTH CLAIMS DATA

Health claims databases have strengths and limitations, depending on the research question that investigators would like to address [71].

Opportunities

Claims databases have several specific advantages over other types of research data, regarding scope, flexibility, costs and statistical power [71].

Regarding the scope of health claims databases, they are generally very comprehensive and complete and often cover the inhabitants of an entire country or region. These databases usually contain data on demographics (e.g. age, sex and postal code), prescribed medication, diagnosis, hospital care and other delivered healthcare, including a complete care pathway of a patient from the first contact with a nephrologist until the last treatment. Since health claims data are collected for payment purposes, they contain data on healthcare costs. Moreover, with health claims data, it is possible to study CKD patients treated with and without RRT in the same data set. Health claims databases often contain data over longer periods of time as well as geographic data, making it possible to study regional differences. Furthermore, since health claims data are collected on a routine basis, they provide information on a non-experimental setting.

The flexibility of health claims data provides opportunities to use different study designs (e.g. cohort studies and case-control design). In addition, for investigators, the data collection is relatively inexpensive and less time-consuming compared with other data collections such as randomized controlled trials or cohort studies. Considering this, with health claims data it is relatively easy to obtain a sufficient number of cases and provide adequate statistical power at relatively low cost [71].

These unique features of health claims databases make it possible to monitor trends in disease prevalence, treatment or healthcare costs over time, providing insight into the effect of changes in policy or guidelines. Therefore health claims data could play a valuable role in guiding health policy and improving quality of care, with the Alberta Kidney Disease Network as an example of a unique collaboration between researchers and policymakers [4]. In addition, health claims data could potentially be used as a quality indicator without providing extra administrative burden for the caregiver, as outcomes can be traced back to individual healthcare providers. For instance, the number of cardiovascular complications in CKD patients, identified using health claims data, can be used to comment on the quality of cardiovascular care on the condition that one can adjust for patient case mix.

Challenges

Health claims data were not designed for clinical research and therefore the researcher cannot control the design, collection and processing of data [16]. Studies have shown that administrative databases, including health claims databases, have limitations in scope (availability of relevant data), data quality and the ability to adjust for patient case mix [72]. Since the majority of health claims databases lack both clinical and laboratory data, the identification of patients is based on specific diagnosis or procedure codes. As a result, the identification of patient groups with health claims data may result in undercoding or overcoding of diagnoses or outcomes [58, 73]. In a database designed for reimbursement purposes, this undercoding or overcoding can be related to coding optimization (i.e. a diagnosis or procedure with higher reimbursement fees is more likely to be coded than the one with lower reimbursement fees). In addition, as previously discussed in the CKD validation studies, usually health claims data are only able to identify CKD patients who are treated by nephrologists while patients who do not come to the attention of health services remain undetected with health claims data. This may be an important limitation in countries that do not have universal healthcare coverage.

Furthermore, the lack of clinical information like severity and progression of the disease, clinical parameters (e.g. smoking and bodyweight) or risk factors might lead, for some research questions, to an incomplete adjustment for potential confounders [71]. Next, in countries without universal healthcare coverage, such as the USA, elderly individuals or those with lower socio-economic status, may be overrepresented in the population [71].

There are several ways to deal with these challenges of health claims data. First, part of the lacking information with regard to the prevalence of chronic diseases or morbidity in CKD patients may be derived from specific medication use. Studies show that data on prescribed medication can be used as a proxy for the prevalence of several chronic diseases [74]. A proxy can be very valid, for example, in case of identifying patients with diabetes mellitus using anti-diabetic drugs and insulin analogues, but less valid if drugs have overlapping indications (e.g. inhalation therapy prescribed for patients with asthma and those with chronic obstructive pulmonary disease).

Second, record linkage of other unrelated administrative databases to health claims databases is a promising tool to add value to the data and it improves and broadens their usage for health research [16]. Record linkage, by means of a unique and direct identifier (i.e. social security number or NHS number), is increasingly used worldwide to combine administrative datasets (e.g. the Alberta kidney disease database) [4]. In case this unique identifier is lacking, indirect linkage tools could be a valid option, although it may introduce less precise record matching [75]. In many countries, maintainers of administrative health databases, such as governmental agencies and health insurance companies, are often reluctant to share administrative databases. The general data protection rules, among others, underlie the restrictive sharing and linking of health data. Therefore the World Health Organization advocates the development of a metadata standard to improve data-sharing policies, thereby increasing the research potential of routinely collected health datasets [76]. We stress the importance of improving the utility of claims data while protecting confidentiality.

Nevertheless, when comparing the results of different health claims databases one must bear in mind the differences in characteristics of the study populations, the regional differences in insurance coverage and the registration of diagnosis or health care use. This might limit the extrapolation of the results to other countries. Apart from important privacy issues this also limits the possibilities of merging international databases. In addition, coding and coverage of diagnoses, procedures or treatments may change over time. One should be aware of these possible changes within a healthcare system when comparing health claims data over a longer period of time.

CONCLUSION

Health claims databases offer important opportunities for studies on large populations of patients with (kidney) disease and health outcomes in a non-experimental setting. However, one should take into account the limitations of health claims data and consider the characteristics of a health claims database, especially when making international comparisons. Since research with health claims data uses codes to identify kidney disease patients and to define other key study variables, information on the validity of these codes in measuring the association of the code with the real variable is indispensable. Available studies indicate that identifying patients undergoing dialysis and the number of performed kidney transplantations using health claims data is accurate, whereas health claims data have low sensitivity for the estimation of the overall CKD prevalence in the general population.

So far, health claims data in 10 countries have been used for studies on kidney disease patients. The unique features of health claims data provide specific research opportunities, such as studying healthcare costs or studying longitudinal, epidemiological data with nationwide coverage. For the optimal utility of health claims data, it is important to keep financial and technical barriers low, while protecting confidentiality. In addition, health claims data can be used to create a nationwide atlas (e.g. the Dutch Kidney Atlas) providing national and regional information on, for instance, the numbers, healthcare costs, prescribed medications, treatments and outcomes of kidney patients.

FUNDING

This work is financed by a grant from the Dutch Kidney Foundation.

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

1. Goldstein BA, Winkelmayer WC. Comparative health services research across populations: the unused opportunities in big data. *Kidney Int* 2015; 87: 1094–1096
2. Kreis K, Neubauer S, Klora M *et al*. Status and perspectives of claims data analyses in Germany—a systematic review. *Health Policy* 2016; 120: 213–226
3. Ng JYS, Ramadani RV, Hendrawan D *et al*. National health insurance databases in Indonesia, Vietnam and the Philippines. *Pharmacoecon Open* 2019; 3: 517–526
4. Bello A, Hemmelgarn B, Manns B *et al*. Use of administrative databases for health-care planning in CKD. *Nephrol Dial Transplant* 2012; 27: iii12–iii18
5. Bello AK, Alrukhaimi M, Ashuntantang GE *et al*. Global overview of health systems oversight and financing for kidney care. *Kidney Int Suppl* 2018; 8: 41–51
6. van Panhuis WG, Paul P, Emerson C *et al*. A systematic review of barriers to data sharing in public health. *BMC Public Health* 2014; 14: 1144
7. Chen Y-C, Yeh H-Y, Wu J-C *et al*. Taiwan's National Health Insurance Research Database: administrative health care database as study object in bibliometrics. *Scientometrics* 2011; 86: 365–380
8. Milea D, Azmi S, Reginald P *et al*. A review of accessibility of administrative healthcare databases in the Asia-Pacific region. *J Mark Access Health Policy* 2015; 3: 28076
9. National Database of Health Insurance Claims and Specific Health Checkups of Japan (NDB). https://www.nttdata.com/global/en/-/media/nttdataglobal/1_files/success-stories/2019/amed.pdf. (6 April 2020, date last accessed)
10. 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–S167
11. Vektis. www.vektis.nl (3 March 2020, date last accessed)

12. Access to ICES (Institute for Clinical Evaluative Sciences) data. https://www.ices.on.ca/DAS/Public-Sector (6 April 2020, date last accessed)
13. Users, Uses and Access to Hospital Episode Statistics. https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/users-uses-and-access-to-hospital-episode-statistics#access-to-data (6 April 2020, date last accessed)
14. 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
15. van Oosten MJM, Logtenberg SJJ, Leegte MJH *et al*. Age-related difference in health care use and costs of patients with chronic kidney disease and matched controls: analysis of Dutch health care claims data. *Nephrol Dial Transplant* 2019
16. Slobbe LC. Working with administrative health data – finding solid ground in the data morass. Thesis, University of Tilburg, Tilburg, 12 June 2019
17. Chang Y-K, Hsu C-C, Chen P-C *et al*. Trends of cost and mortality of patients on haemodialysis with end stage renal disease. *Nephrology (Carlton)* 2015; 20: 243–249
18. Chang Y-T, Hwang J-S, Hung S-Y *et al*. Cost-effectiveness of hemodialysis and peritoneal dialysis: a national cohort study with 14 years follow-up and matched for comorbidities and propensity score. *Sci Rep* 2016; 6: 30266
19. Couchoud C, Couillerot A-L, Dantony E *et al*. Economic impact of a modification of the treatment trajectories of patients with end-stage renal disease. *Nephrol Dial Transplant* 2015; 30: 2054–2068
20. Couillerot-Peyrondet A-L, Sambuc C, Sainsaulieu Y *et al*. A comprehensive approach to assess the costs of renal replacement therapy for end-stage renal disease in France: the importance of age, diabetes status, and clinical events. *Eur J Health Econ* 2017; 18: 459–469
21. Helmuth ME, Liu Q, Turenne MN *et al*. Secular trends in the cost of immunosuppressants after solid organ transplantation in the United States. *Clin J Am Soc Nephrol* 2019; 14: 421–430
22. Honeycutt AA, Segel JE, Zhuo X *et al*. Medical costs of CKD in the Medicare population. *J Am Soc Nephrol* 2013; 24: 1478–1483
23. Kao T-W, Chang Y-Y, Chen P-C *et al*. Lifetime costs for peritoneal dialysis and hemodialysis in patients in Taiwan. *Perit Dial Int* 2013; 33: 671–678
24. Kitazawa T, Matsumoto K, Fujita S *et al*. Cost analysis of transplantation in Japan, performed with the use of the national database. *Transplant Proc* 2017; 49: 4–9
25. 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
26. Shaikh M, Woodward M, John O *et al*. Utilization, costs, and outcomes for patients receiving publicly funded hemodialysis in India. *Kidney Int* 2018; 94: 440–445
27. Chettiar A, Montez-Rath M, Liu S *et al*. Association of inpatient palliative care with health care utilization and postdischarge outcomes among Medicare beneficiaries with end stage kidney disease. *Clin J Am Soc Nephrol* 2018; 13: 1180–1187
28. Choi Y, Shin J, Park JT *et al*. Disparities in kidney transplantation access among Korean patients initiating dialysis: a population-based cohort study using National Health Insurance data (2003–2013.). *Am J Nephrol* 2017; 45: 32–39
29. Dobbels F, Skeans MA, Snyder JJ *et al*. Depressive disorder in renal transplantation: an analysis of Medicare claims. *Am J Kidney Dis* 2008; 51: 819–828
30. Farrugia D, Mahboob S, Cheshire J *et al*. Malignancy-related mortality following kidney transplantation is common. *Kidney Int* 2014; 85: 1395–1403
31. Ferreira JP, Couchoud C, Gregson J *et al*. Angiotensin-converting enzyme inhibitors/angiotensin receptor blockers, β-blockers or both in incident end-stage renal disease patients without cardiovascular disease: a propensity-matched longitudinal cohort study. *Nephrol Dial Transplant* 2019; 34: 1216–1222
32. Ferro CJ, Arnold J, Bagnall D *et al*. Fracture risk and mortality post-kidney transplantation. *Clin Transplant* 2015; 29: 1004–1012
33. Han SS, Park JY, Kang S *et al*. Dialysis modality and mortality in the elderly: a meta-analysis. *Clin J Am Soc Nephrol* 2015; 10: 983–993
34. Hayer MK, Farrugia D, Begaj I *et al*. Infection-related mortality is higher for kidney allograft recipients with pretransplant diabetes mellitus. *Diabetologia* 2014; 57: 554–561
35. Hung S-C, Chang Y-K, Liu J-S *et al*. Metformin use and mortality in patients with advanced chronic kidney disease: national, retrospective, observational, cohort study. *Lancet* 2015; 3: 605–614
36. Kim H, Kim KH, Ahn SV *et al*. Risk of major cardiovascular events among incident dialysis patients: a Korean national population-based study. *Int J Cardiol* 2015; 198: 95–101
37. Kitchlu A, Clemens K, Gomes T *et al*. Beta-blockers and cardiovascular outcomes in dialysis patients: a cohort study in Ontario, Canada. *Nephrol Dial Transplant* 2012; 27: 1591–1598
38. Komenda P, Yu N, Leung S *et al*. Secular trends in end-stage renal disease requiring dialysis in Manitoba, Canada: a population-based study. *CMAJ Open* 2015; 3: E8–E14
39. Kuo H-W, Tsai S-S, Tiao M-M *et al*. Epidemiological features of CKD in Taiwan. *Am J Kidney Dis* 2007; 49: 46–55
40. Lam NN, Kim SJ, Knoll GA *et al*. The risk of cardiovascular disease is not increasing over time despite aging and higher comorbidity burden of kidney transplant recipients. *Transplantation* 2017; 101: 588–596
41. Lenihan CR, Liu S, Montez-Rath ME *et al*. Trends in the medical complexity and outcomes of Medicare-insured patients undergoing kidney transplant in the years 1998–2014. *Transplantation* 2019; 11: 2413–2422
42. Li W-H, Chen Y-J, Tseng W-C *et al*. Malignancies after renal transplantation in Taiwan: a nationwide population-based study. *Nephrol Dial Transplant* 2012; 27: 833–839
43. Liao J-N, Chao T-F, Liu C-J *et al*. Incidence and risk factors for new-onset atrial fibrillation among patients with end-stage renal disease undergoing renal replacement therapy. *Kidney Int* 2015; 87: 1209–1215
44. Rene E, Lazrak HH, Laurin L-P *et al*. Association of erythropoiesis-stimulating agents and the incidence risk of cancer diagnosis among chronic dialysis patients: a nested case-control study. *Nephrol Dial Transplant* 2017; 32: 1047–1052
45. Tonelli M, Wiebe N, James MT *et al*. A population-based cohort study defines prognoses in severe chronic kidney disease. *Kidney Int* 2018; 93: 1217–1226
46. Wang F, Yang C, Long J *et al*. Executive summary for the 2015 Annual Data Report of the China Kidney Disease Network (CK-NET). *Kidney Int* 2019; 95: 501–505
47. Wang H-H, Hung S-Y, Sung J-M *et al*. Risk of stroke in long-term dialysis patients compared with the general population. *Am J Kidney Dis* 2014; 63: 604–611

48. Wang I-K, Cheng Y-K, Lin C-L *et al*. Comparison of subdural hematoma risk between hemodialysis and peritoneal dialysis patients with ESRD. *Clin J Am Soc Nephrol* 2015; 10: 994–1001
49. Wang I-K, Lin C-L, Chen H-C *et al*. Risk of new-onset diabetes in end-stage renal disease patients undergoing dialysis: analysis from registry data of Taiwan. *Nephrol Dial Transplant* 2018; 33: 670–675
50. Weinhandl ED, Ray D, Kubisiak KM *et al*. Contemporary trends in clinical outcomes among dialysis patients with Medicare coverage. *Am J Nephrol* 2019; 50: 63–71
51. Wu C-Y, Wu M-S, Kuo KN *et al*. Long-term peptic ulcer rebleeding risk estimation in patients undergoing haemodialysis: a 10-year nationwide cohort study. *Gut* 2011; 60: 1038–1042
52. Wu C-C, Chen S-H, Ho C-H *et al*. End-stage renal disease after hypertensive disorders in pregnancy. *Am J Obstet Gynecol* 2014; 210: 147.e1–147.e8
53. Yoon C-Y, Noh J, Jhee JH *et al*. Warfarin use in patients with atrial fibrillation undergoing hemodialysis: a nationwide population-based study. *Stroke* 2017; 48: 2472–2479
54. Yu T-M, Chuang Y-W, Yu M-C *et al*. Risk of cancer in patients with polycystic kidney disease: a propensity-score matched analysis of a nationwide, population-based cohort study. *Lancet Oncol* 2016; 17: 1419–1425
55. 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
56. 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
57. Ronksley PE, Tonelli M, Quan H *et al*. Validating a case definition for chronic kidney disease using administrative data. *Nephrol Dial Transplant* 2012; 27: 1826–1831
58. 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
59. van Oosten M, Brohet R, Logtenberg S *et al*. FP414 The validity of Dutch health claims data in identifying patients with chronic kidney disease. *Nephrol Dial Transplant* 2019; 34(Suppl 1):
60. Clement FM, James MT, Chin R *et al*. Validation of a case definition to define chronic dialysis using outpatient administrative data. *BMC Med Res Methodol* 2011; 11: 25
61. 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
62. 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
63. Lam NN, McArthur E, Kim SJ *et al*. Validation of kidney transplantation using administrative data. *Can J Kidney Health Dis* 2015; 2: 54
64. Chen Y-C, Wu J-C, Chen T-J *et al*. Reduced access to database. A publicly available database accelerates academic production. *BMJ* 2011; 342: d637
65. Dutch Health Insurance Act, 2016. http://wetten.overheid.nl/BWBR0018450/2016-01-01
66. Organisation for Economic Co-operation and Development (OECD). *Social Protection*. https://stats.oecd.org/viewhtml.aspx?datasetcode=HEALTH_PROT&lang=en
67. Centraal Bureau voor de Statistiek (CBS). *Bevolking; kerncijfers*, 2017. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/37296ned/table?fromstatweb
68. 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
69. van Walraven C, Bennett C, Forster AJ. Administrative database research infrequently used validated diagnostic or procedural codes. *J Clin Epidemiol* 2011; 64: 1054–1059
70. Akobeng AK. Understanding diagnostic tests 1: sensitivity, specificity and predictive values. *Acta Paediatr* 2007; 96: 338–341
71. Motheral BR, Fairman KA. The use of claims databases for outcomes research: rationale, challenges, and strategies. *Clin Ther* 1997; 19: 346–366
72. Mohammed MA, Stevens A. The value of administrative databases. *BMJ* 2007; 334: 1014–1015
73. 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
74. Slobbe LCJ, Fussenich K, Wong A *et al*. Estimating disease prevalence from drug utilization data using the Random Forest algorithm. *Eur J Public Health* 2019; 29: 615–621
75. Raffray M, Bayat S, Lassalle M *et al*. Linking disease registries and nationwide healthcare administrative databases: the French Renal Epidemiology and Information Network (REIN) insight. *BMC Nephrol* 2020; 21: 25
76. Pisani E, AbouZahr C. Sharing health data: good intentions are not enough. *Bull World Health Organ* 2010; 88: 462–466