

Developing and validating an algorithm to identify incident chronic dialysis patients using administrative data

Dino Gibertoni¹, Claudio Voci², Marica Iommi³, Benedetta D'Ercole⁴, Marcora Mandreoli⁵, Antonio Santoro⁶, Elena Mancini⁷

¹ Department of Biomedical and Neuromotor Sciences, University of Bologna, Italy

² Bologna Local Health Authority, Bologna, Italy

³ Advanced School for Health Policy, University of Bologna, Italy

⁴ Cooperativa EDP La Traccia, Matera, Italy

⁵ Nephrology and Dialysis Unit, S. Maria della Scaletta Hospital, Imola, Italy

⁶ Specialty School in Nephrology, University of Bologna, Italy

⁷ Nephrology, Dialysis and Hypertension Unit, Policlinico S.Orsola-Malpighi, Bologna, Italy

Corresponding author: Dino Gibertoni, Department of Biomedical and Neuromotor Sciences, University of Bologna, Via San Giacomo 12, 40126 Bologna Italy. e-mail: dino.gibertoni2@unibo.it

Abstract

Background. Administrative healthcare databases are widespread and are often standardized with regard to their content and data coding, thus they can be used also as data sources for surveillance and epidemiological research. Chronic dialysis requires patients to frequently access hospital and clinic services, causing a heavy burden to healthcare providers. This also means that these patients are routinely tracked on administrative databases, yet very few case definitions for their identification are currently available. The aim of this study was to develop two algorithms derived from administrative data for identifying incident chronic dialysis patients and test their validity compared to the reference standard of the regional dialysis registry.

Methods. The algorithms are based on data retrieved from hospital discharge records (HDR) and ambulatory specialty visits (ASV) to identify incident chronic dialysis patients in an Italian region. Subjects are included if they have at least one event in the HDR or ASV databases based on the ICD9-CM dialysis-related diagnosis or procedure codes in the study period. Exclusion criteria comprise non-residents, prevalent cases, or patients undergoing temporary dialysis, and are evaluated only on ASV data by the first algorithm, on both ASV and HDR data by the second algorithm. We validated the algorithms against the Emilia-Romagna regional dialysis registry by searching for incident patients in 2014.

Results. Algorithm 1 identified 680 patients and algorithm 2 identified 676 initiating dialysis in 2014, compared to 625 patients included in the regional dialysis registry. Sensitivity for the two algorithms was respectively 90.8% and 88.4%, positive predictive value 84.0% and 82.0%, and percentage agreement was 77.4% and 74.1%.

Conclusions. These results suggest that administrative data have high sensitivity and positive predictive value for the identification of incident chronic dialysis patients. Algorithm 1, which showed the higher accuracy and has a simpler case definition, can be used in place of regional dialysis registries when they are not present or sufficiently developed in a region, or to improve the accuracy and timeliness of existing registries.

Keywords

Chronic dialysis, administrative data, hospital discharge records, ambulatory specialty visits, case definition, algorithm

Background

The prevalence of end stage kidney disease (ESKD) requiring dialytic treatment is growing worldwide, due to population ageing and the increased prevalence of comorbidities (1–3). Patients on dialysis have a high mortality risk (4–6) and their management is crucial for health care providers, because of the high organizational and financial burden related to the frequency and the complexity of the treatment (7–10). Therefore, it is extremely important to implement population-based registries of patients in dialysis, to assist health care providers in organizing the treatment and to have data available to evaluate the patients' health outcomes. Many national or regional registries of patients on dialysis or on renal replacement therapy (RRT) have been created since the 1990s; as a result, in Europe the ERA-EDTA Registry currently gathers information from more than 50 countries and regions and produces yearly a detailed report (11). However, to implement and regularly maintain a regional dialysis registry a large amount of dedicated human resources would be needed. This is actually true in Italy, that conferred data from only 8 out of 21 regions for the last published ERA-EDTA registry report (11). One way to overcome this problem is to exploit the data recorded in the official administrative databases and create a regional registry by means of an automated algorithm. Very few algorithms and case definitions for incident dialysis patients can be found in the literature: Clement et al (12) presented and validated a case definition based on Canadian outpatient physician billings, and other algorithms to identify RRT patients from administrative data were proposed using data from Australian and Italian regions (13,14). In this study, we submit a novel algorithm that identifies incident dialysis patients taking advantage of data from the hospital discharges and outpatient specialty databases. We validated the algorithm by comparison with the gold standard represented by the dialysis registry of the Emilia-Romagna region. This algorithm relies on ICD9-CM coding and can be easily adapted to be utilized in other region and countries, where it could aid to implement a new registry of chronic dialysis patients or improve the quality of existing registries.

Methods

Algorithms

We describe two algorithms designed to identify the incident chronic dialysis patients in a selected time period (the *index period*) using data obtained from the regional hospital discharge records (HDR) and ambulatory specialty visits (ASV) databases. The algorithms require all dialysis events recorded from 1 year before the starting date to 1 year after the ending date of the index period. In this study, we tested the algorithms by identifying the incident chronic dialysis patients in 2014,

thus we built a dataset containing hospital admissions and ambulatory specialty visits from 1.1.2013 to 31.12.2015. The two algorithms used the same inclusion criteria:

- in HDR, at least one admission in the index period with ICD9-CM main or secondary diagnosis codes 585.6 (end stage renal disease, excluding admissions aimed to create the arteriovenous or peritoneal fistula, identified by 39.27, 39.29 or 54.93 codes as the only dialysis-related procedure code), V45.1 (renal dialysis status), or V56.0-V56.8 (dialysis encounter). In this case, the admission date was used to set the dialysis date;
- or, in HDR, at least one admission with any procedure code ICD9-CM 39.95 (hemodialysis) or 54.98 (peritoneal dialysis) In this case, the date of the procedure was used to set the dialysis date; when more than one dialysis procedure was found, the earliest date was used. If the HDR record included both diagnosis and procedure codes related to dialysis, the procedure date was selected;
- or, in ASV, at least one hemodialysis (codes 39.95.1-39.95.9) or at least one peritoneal dialysis (codes 54.98.1 or 54.98.2) in the index year.

For each patient, the date of the first dialysis (*index date*) was defined as the earliest among the dates of visits and hospital admissions of 2014.

Exclusion criteria were:

- patients not residing in the Emilia-Romagna region: they were excluded if at least one record from the HDR and ASV databases in the index period indicated a different region of residence;
- prevalent cases: patients were excluded if at least one hemodialysis or peritoneal dialysis was found in the 365 days before the index date;
- non-chronic dialysis patients: patients were excluded if, within one year following the index date, less than 30 days passed between the index date and the last dialysis date, or less than 90 days for patients who initiated dialysis for acute kidney injury (AKI, primary or secondary diagnosis code ICD9-CM 584.x found in the index dialysis admission or in admissions occurred less than 90 days before the index date). This criterion allows to exclude patients who died, or recovered, or were transferred to a nephrology unit of another region shortly after dialysis initiation;
- patients who died during the hospitalization in which the first dialytic treatment was provided. Only algorithm 2 used this criterion.

The two algorithms search for the index dialysis in both HDR and ASV databases, but differ in the definition of chronic dialysis: algorithm 1 defines as chronic dialysis patients those who are seen regularly in outpatient clinics, and therefore searches for previous and subsequent dialysis events

only in the ASV database. Algorithm 2 considers also dialysis treatments provided during hospitalizations, and thus uses data from both the ASV and the HDR databases. The algorithms were developed in Stata 15.1 and SAS Enterprise Guide 7.1.

Accuracy of algorithms

The algorithms were tested on data taken from the administrative databases of Emilia-Romagna, a region in north-western Italy with 11 local health authorities in 2014 (currently merged into 8) serving a population of 4.446 million inhabitants as of 1/1/2014 (data provided by Istat at <https://www.istat.it/en/population-and-households?data-and-indicators>). According to the regional dialysis registry, in 2014 the number of incident dialysis patients in Emilia-Romagna was 685, corresponding to 154 per million population (pmp), compared to 159 pmp in Italy. (15)

Records from the Emilia-Romagna regional administrative databases were linked using a unique pseudonymized patient identifier, which allows capturing all patient encounters with the health care system. The data taken from the HDR and ASV databases to feed the algorithms belong to a subset of compulsory data homogeneously recorded in all Italian regions, as they are an information obligation to the Ministry of Health. Specifically, the hospital discharge records database contains admission date, discharge date, intervention date, specialty at discharge, up to 15 diagnostic codes and 15 procedure codes [International Classification of Disease (ICD-9 CM)] for each admission. The ASV database contains records regarding each hemodialysis session or one monthly summary record for peritoneal dialysis; in both cases, data on the type of service delivered (laboratory test, specialty service, rehabilitation service), the medical discipline related to the service, the date of delivery are present. Due to the current Italian regulation on privacy, we could not use the regional patient registry to obtain patients' personal data, thus the residence of patients was ascertained from the data present in the HDR and ASV databases.

Gold standard

The Emilia-Romagna regional Dialysis Registry (ERDR) was used as the reference source to evaluate the accuracy of the two algorithms. The ERDR was established in 1994 and is part of the Italian Dialysis Registry network, which in turn feeds the ERA-EDTA Dialysis Registry; as such it is the official source of epidemiological information about chronic dialysis in the Emilia-Romagna region. Incident cases in the ERDR are patients living in the Emilia-Romagna region who initiate dialysis for the first time according to the intention-to-treat approach. The dialysis inception date reported in the ERDR is assigned by the nephrologist in charge, and is the date in which the patient started being considered as chronic. Patients who stop dialysis because of transplantation are

maintained in the registry as “transplanted patients” (hence, they are not included in the population of prevalent dialysis patients) and if they subsequently reinstate dialysis, they will be considered as new entries for that year. Patients starting dialysis following an AKI and not chronicised, as well as guests (patients temporarily treated with dialysis in a regional clinic but living elsewhere, in Italy or abroad) are not added to the ERDR. The registry is updated yearly by transferring data from the information systems of nephrology units, checked for data quality by the nephrologist in charge of the registry (EM) and uploaded on a dedicated website (<https://www.regdial.it>). In 2014 all but one of the local nephrology units could automatically transfer data from their patients’ management information systems to the ERDR, therefore widely reducing the possible mistakes due to manual data transcription. However, the ERDR is regularly monitored and revised for inconsistencies and missing data.

The comparison between the algorithms and the ERDR was made by calculating only percentage agreement, sensitivity and positive predictive value (PPV), because the ERDR does not contain subjects without dialysis.

Results

Overall 610,724 records of 6,108 patients with at least one dialysis event in 2014 were retrieved from HDR and ASV. From these, data provided by two local health authorities were excluded from the analysis because their ASV records data for the years 2013-14 were largely incomplete. After applying exclusion criteria, 680 incident chronic dialysis patients were identified by algorithm 1 and 676 by algorithm 2. The patients mutually identified by both algorithms were 631. The ERDR included 625 incident patients in 2014 after removing those treated in the two local health authorities with incomplete data. Algorithm 1 showed sensitivity=90.8% and PPV=84.0% (Table 1) and algorithm 2 had a slightly lower performance (sensitivity =88.5% and PPV=82.0%). Algorithm 2 was less accurate mostly because it identified more chronic incident in dialysis patients unknown to ERDR (68 vs. 39) on the basis of occasional hospital admissions while not having had any dialysis related specialty visit in the year following the index date. The estimated incidence rate of dialysis inception per million regional population was 168 for ERDR, 182 for algorithm 1 and 181 for algorithm 2.

The socio-demographic and clinical characteristics of the cohorts identified by the algorithms and those included in ERDR were overall very similar (Table 2): algorithm 1’s cohort was slightly younger than the ERDR’s cohort, and as such had a lower HD type of dialysis, lower number of treatments and less comorbidities.

The main reasons for lack of agreement between the algorithms and ERDR were: a) discordant information on patients' region of residence; b) patients recorded with different index dates, falling in two adjacent calendar years. For instance, a patient could be identified as prevalent by the algorithm on the basis of the treatment he received during a hospital admission in December 2013, while in the ERDR it was recorded as incident in January 2014; c) patients included in the ERDR although they were treated for less than 30 days (or 90 days in the case of AKI at the index dialysis), because of death, functional recovery or relocation to another Italian region or abroad.

Discussion

The two algorithms developed in this paper showed good accuracy in the identification of incident chronic dialysis patients from routinely collected administrative data. Both algorithms are easy to implement because they need as input data taken from the hospital discharge records and outpatient specialty visits databases, which are implemented in many countries, and apply inclusion and exclusion criteria based on ICD-9 CM codes. Algorithm 1 can be preferred because it displayed the highest accuracy compared to the dialysis registry (90.8% sensitivity, 84.0% PPV) and it is based on a simpler case definition.

The accuracy of our algorithm is consistent to the findings of other case definitions of chronic dialysis that were elaborated in Canada (12,16). As these authors pointed out, the accuracy of their case definitions increased as the time interval on which they were applied was extended from 1 year to 5 years, thus we deem that the accuracy of our algorithm might also improve when searching for incident patients on a multi-year period. Actually, one of the reasons of mismatch between our algorithms and the ERDR reference is the date of dialysis inception recorded in those sources for the same patient, which may differ slightly and fall in two different calendar years. Clearly, this kind of discrepancy becomes less influent as the time period for incident patients' identification is extended. Another relevant reason of mismatch was the inconsistency between the reported regions of residence. In regions with high migration rates there could actually be a non-negligible number of patients who change their residence near the dialysis inception and whose residency may be erroneously coded. However, this bias may be somewhat inflated in our validation, because we obtained data on residence from the ASV and HDR records which are less accurate than the regional registry of patients. There are also patients who experienced a complex therapeutic trajectory, alternating periods of dialysis treatment and periods of conservative treatment, which can make it difficult for an automated algorithm to discriminate whether they are under chronic or temporary treatment.

One peculiar feature of our algorithm is having established at 30 days after the index date the time needed to determine whether a patient initiating dialysis is chronic, in contrast to the more conventional minimum duration of 90 days (17). We think that patients who underwent treatment for 30 to 90 days are worthy being considered as receiving chronic dialysis treatment, as they needed a relevant amount of resources, generated a non-negligible cost to the institution in which they were treated and may experience a renal outcome in that timeframe. Thus, we deem our case definition is more accurate to identify cohorts of patients to be considered in studies aimed to estimate the costs of dialysis treatment and, more generally, for epidemiological studies and outcome studies, taking advantage of the possibility to link the patients' records to the other administrative databases.

An efficient case definition algorithm may also be useful to automatically feed a regional registry of dialysis patients, especially in those regions in which a registry is not yet established. Data from administrative databases have the advantage over manually fed data of being population-based and less error-prone (18), provided of course that the data source is of good quality. The main limitation of an automated algorithm lies indeed in its dependence on the scope and quality of the administrative data, that are usually collected for purposes different from research. Thus, the implementation of an algorithm might also be a trigger for the identification of inaccuracy areas in the administrative data and lead to intervention aimed to improve the quality of data. Another limitation is the difficulty to correctly identify complex cases that do not fall within the criteria defined by the algorithm and require the experience of nephrologists to be adequately classified.

Conclusions

We provide two algorithms that identify with good accuracy incident chronic dialysis patients from the administrative databases of hospital discharges and outpatient visits. The algorithm that identifies patients as chronic dialysis incidents if they attended ambulatory specialty visits after the first dialysis is to be preferred, because it is more accurate. These algorithms may be useful to create regional registries of chronic dialysis patients or to improve the accuracy of existing registries. As they are derived from administrative databases using ICD9-CM codes, with the necessary adaptations they can be used in many regional or national settings and can easily be linked to other data sources, representing a valuable tool for clinical and epidemiological studies. The performance of our algorithms should be further validated using data from other regions.

Abbreviations

HDR: Hospital discharge records; ASV: Ambulatory specialty visits; ERDR: Emilia-Romagna dialysis registry; ESKD: End stage kidney disease; RRT: renal replacement therapy; AKI: acute kidney injury

Declarations

Ethics approval and consent to participate

The study was exempt from approval from the Ethics Committee of Emilia-Romagna. It was conducted in conformity with the regulations for data management from the Regional Health Authority of Emilia-Romagna, and with the Italian Code of conduct and professional practice applying to processing of personal data for statistical and scientific purposes (art. 20–21, legislative decree 196/2003;

<https://www.garanteprivacy.it/documents/10160/0/Codice+in+materia+di+protezione+dei+dati+personali+%28Testo+coordinato%29>) published in the Official Journal no. 190 of August 14, 2004, which explicitly exempts the need for approval from the Ethics Committee when using anonymous data (Preamble number 8).

In Italy, anonymous administrative data-gathering is subject to the law Protection of individuals and other subjects with regard to the processing of personal data, Act no. 675 of 31.12.1996 (<https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/28335>, amended by Legislative Decree no. 123 of 09.05.1997, no. 255 of 28.07.1997, no. 135 of 08.05.1998, no.171 of 13.05.1998, no. 389 of 6.11.1998, no. 51 of 26.02.1999, no. 135 of 11.05.1999, no. 281 of 30.07.1999, no.282 of 30.07.1999 and no. 467 of 28.12.2001). Patients enrolled in the Emilia-Romagna Dialysis Registry provided written consent to use their anonymized data.

Consent for publication

Not applicable

Availability of data and materials

The datasets generated and/or analysed during the current study are not publicly available due to the current regulation on privacy but are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests

Funding

No funding was received.

Authors' contributions

The research idea was conceived by DG and CV. The algorithms were defined by EM, MM, DG, MI, and the SAS and Stata scripts were written by CV, MI and DG. CV and BDE provided and extracted the data. DG, AS and MM wrote the first draft of the manuscript and all six authors (DG, CV, MI, BDE, MM, AS and EM) contributed to the review of the manuscript, provided critical comments and approved the final version.

Acknowledgments

The authors wish to thank the nephrologists which contributed to the Emilia-Romagna Dialysis Registry: Sara De Amicis (Ospedale "Guglielmo da Saliceto", Piacenza), Gabriele Donati (Policlinico S.Orsola-Malpighi, Bologna), Marta Flachi (Ospedale Infermi, Rimini), Davide Gerra, (Parma Local Health Authority), Carmela Giovannone (Ospedale "B. Ramazzini", Carpi), Antonio Giudicissi (Ospedale "M.Bufalini", Cesena), Elisabetta Isola (Ospedale "S.Maria delle Croci", Ravenna), Leonardo Lucchi (Policlinico di Modena), Lucio Manenti (Ospedale Maggiore, Parma), Renato Rapanà (Ospedale "S.Maria della Scaletta", Imola), Giorgia Russo (Arcispedale "S.Anna", Ferrara), Alfredo Stefani (Arcispedale "S.Maria Nuova", Reggio Emilia), Loretta Zambianchi (Ospedale "G.B.Morgagni-L.Pierantoni", Forlì)

References

1. Go AS, Chertow GM, Fan D, Mcculloch CE, Hsu C. Chronic Kidney Disease and the Risks of Death, Cardiovascular Events, and Hospitalization. *N Engl J Med.* 2004;13(351):1296–305.
2. De Lissovoy G. Measuring the burden of illness for end-stage renal disease: some heavy lifting required. *Kidney Int.* 2007;72(9):1045–7.
3. Tonelli M, Wiebe N, Guthrie B, James MT, Quan H, Fortin M, et al. Comorbidity as a driver of adverse outcomes in people with chronic kidney disease. *Kidney Int.* 2015;88(4):859–66.
4. Tonelli M, Wiebe N, Culeton B, House A, Rabbat C, Fok M, et al. Chronic kidney disease and mortality risk: a systematic review. *J Am Soc Nephrol.* 2006 Jul;17(7):2034–47.
5. Nordio M, Limido A, Maggiore U, Nichelatti M, Postorino M, Quintaliani G. Survival in patients treated by long-term dialysis compared with the general population. *Am J Kidney*

Dis. 2012 Jun;59(6):819–28.

6. United States Renal Data System. 2018 USRDS annual data report: Epidemiology of kidney disease in the United States. Bethesda, MD; 2018.
7. Levey AS, Atkins R, Coresh J, Cohen EP, Collins AJ, Eckardt KU, et al. Chronic kidney disease as a global public health problem: Approaches and initiatives - a position statement from Kidney Disease Improving Global Outcomes. *Kidney Int.* 2007;72:247–59.
8. Pontoriero G, Pozzoni P, Del Vecchio L, Locatelli F. International Study of Health Care Organization and Financing for renal replacement therapy in Italy: an evolving reality. *Int J Health Care Finance Econ.* 2007;7(2–3):201–15.
9. Roggeri A, Roggeri DP, Zocchetti C, Bersani M, Conte F. Healthcare costs of the progression of chronic kidney disease and different dialysis techniques estimated through administrative database analysis. *J Nephrol.* 2017 Apr 1;30(2):263–9.
10. Roggeri DP, Roggeri A, Salomone M. Chronic Kidney Disease: Evolution of Healthcare Costs and Resource Consumption from Predialysis to Dialysis in Piedmont Region, Italy. *Adv Nephrol.* 2014;2014:1–6.
11. ERA-EDTA Registry. ERA-EDTA Registry Annual Report 2017. Amsterdam, the Netherlands; 2019.
12. Clement FM, James MT, Chin R, Klarenbach SW, Manns BJ, Quinn RR, et al. Validation of a case definition to define chronic dialysis using outpatient administrative data. *BMC Med Res Methodol.* 2011;11(25).
13. Brameld KJ, Thomas MAB, Holman CDAJ, Bass AJ, Rouse IL. Validation of linked administrative data on end-stage renal failure: Application of record linkage to a “clinical base population.” *Aust N Z J Public Health.* 1999;23(5):464–7.
14. Valent F, Busolin A, Boscutti G. Inception and utility of a renal replacement registry using administrative health data in North-East Italy. *J Ren Care.* 2017 Jun 1;43(2):121–7.
15. RIDT - Registro Italiano di Dialisi e Trapianto. Report 2014 [Internet]. 2017 [cited 2019 Oct 10]. Available from: <https://ridt.sinitaly.org/2017/03/21/report-2014/>
16. Komenda P, Yu N, Leung S, Bernstein K, Blanchard J, Sood M, et al. Determination of the

optimal case definition for the diagnosis of end-stage renal disease from administrative claims data in Manitoba, Canada. *C Open*. 2015;3(3):E8–14.

17. Lok CE, Oliver MJ, Rothwell DM, Hux JE. The growing volume of diabetes-related dialysis: A population based study. *Nephrol Dial Transplant*. 2004;19(12):3098–103.
18. Chowdhury TT, Hemmelgarn B. Evidence-Based Decision-Making 6: Utilization of Administrative Databases for Health Services Research. In: Parfrey PS, Barrett BJ, editors. *Clinical Epidemiology: Practice and Methods*. Second. New York, NY: Springer Science+Business Media, Inc.; 2015. p. 469–84.

Table 1 Accuracy of the two algorithms vs. the Emilia-Romagna Dialysis (ERDR) registry

	Algorithm 1	Algorithm 2
Incident cases identified by the algorithm	680	676
found in ERDR, incident in 2014	571	554
found in ERDR, incident in a different year	70	54
not found in ERDR	39	68
Incident ERDR cases in 2014 missed by the algorithm	58	72
Percentage agreement	77.4%	74.1%
Sensitivity	90.8%	88.5%
Positive predictive value (PPV)	84.0%	82.0%

Table 2 Characteristics of the cohorts identified by algorithms and by the regional dialysis registry

	Algorithm 1	Algorithm 2	ERDR
N	680	676	625
Age, mean±sd	67.6±15.0	67.3±15.3	68.0±14.6
Females, n (%)	233(34.3)	236(34.9)	215(34.7)
Haemodialysis, n (%)	591(86.9)	591(87.4)	546(88.2)
n. of treatments in the first year, median (p25-p75)	24(11-103)	28(11-105)	26(11-106)
n. of hospital admissions in the first year, mean±sd	2.6±2.2	2.7±2.2	2.1±1.7
Charlson Comorbidity Index, median (p25-p75)	1(0-3)	2(1-3)	2(1-3)