Identification of Patients with CKD in Medical Databases
A Comparison of Different Algorithms

Søren Viborg Vestergaard1,2, Christian Fynbo Christiansen1,1, Reimar Wernich Thomsen2,1, Henrik Birn2,3 and Uffe Heide-Jørgensen1

Abstract

Background and objectives Despite CKD consensus definitions, epidemiologic studies use multiple different algorithms to identify CKD. We aimed to elucidate if this affects the patient characteristics and the estimated prevalence and prognosis of CKD by applying six different algorithms to identify CKD in population-based medical databases and compare the cohorts.

Design, setting, participants, & measurements Patients with CKD in Northern Denmark (2009–2016) were identified using six different algorithms: five were laboratory based defined by (1) one measured outpatient eGFR <60 ml/min per 1.73 m² (single test, n = 103,435), (2) two such findings ≥90 days apart (Kidney Disease Improving Global Outcomes, n = 84,688), (3) two such findings ≥90 days apart with no eGFR >60 ml/min per 1.73 m² observed in-between (Kidney Disease Improving Global Outcomes, persistent, n = 68,994), (4) two such findings ≥90 and <365 days apart (Kidney Disease Improving Global Outcomes, time limited, n = 75,031), and (5) two eGFRs <60 ml/min per 1.73 m² or two urine albumin-creatinine ratios >30 mg/g ≥90 days apart (Kidney Disease Improving Global Outcomes, eGFR/albuminuria, n = 100,957). The sixth included patients identified by reported in- and outpatient hospital International Classification of Diseases diagnoses of CKD (hospital-diagnosed, n = 27,947). For each cohort, we estimated baseline eGFR, CKD prevalence, and 1-year mortality using the Kaplan–Meier method.

Results The five different laboratory-based algorithms resulted in large differences in the estimated prevalence of CKD from 4637–8327 per 100,000 population. In contrast, 1-year mortality varied only slightly (7%–9%). Baseline eGFR levels at diagnosis were comparable (53–56 ml/min per 1.73 m²), whereas median time since first recorded eGFR <60 ml/min per 1.73 m² varied from 0 months (single-test) to 17 months (Kidney Disease Improving Global Outcomes, persistent). The hospital-diagnosed algorithm yielded markedly lower CKD prevalence (775 per 100,000 population), a lower baseline eGFR (47 ml/min per 1.73 m²), longer time since first eGFR <60 ml/min per 1.73 m² (median 70 months), and much higher 1-year mortality (22%).

Conclusions Population prevalence of CKD identified in medical databases greatly depends on the applied algorithm to define CKD. Despite these differences, laboratory-based algorithms produce cohorts with similar prognosis.

Introduction

Epidemiologic studies using medical databases play an important role in clarifying the prevalence and outcomes of CKD (1–3). Most patients with mild or moderate CKD are managed by primary care physicians (4), and hospital databases may be incomplete and not correctly identify the time of onset of CKD (5). In addition to structural abnormalities, the Kidney Disease Improving Global Outcomes (KDIGO) guidelines define CKD strictly by laboratory findings, including eGFR and/or abnormal urinary findings, which must be present for more than 3 months (6). Often, structural abnormalities of the kidneys are not known or reported, and thus, clinical biochemistry databases holding information on eGFR and urine albumin-creatinine ratio (UACR) data provide an attractive alternative to hospital databases for more timely identification and accurate staging of patients with CKD. Because albuminuria may not be quantitatively assessed or consistently reported, epidemiologic studies on CKD are often on the basis of eGFR only.

Despite the CKD consensus definitions, the practical application is ambiguous and multiple different algorithms are used to identify CKD in epidemiologic studies (7,8). This may lead to differences in estimated incidence, prevalence, and prognosis of CKD. These differences need to be clarified to enable correct interpretation and comparison of the findings.

To address this issue, we applied six different previously used algorithms to identify patients with CKD in Danish population-based registry data on
laboratory test results or hospital diagnoses. The aim was to characterize the differences between the resulting cohorts with respect to cohort size, baseline characteristics, kidney function, and the estimated prevalence, incidence, and prognosis of CKD.

**Materials and Methods**

We conducted the study in Northern Denmark (corresponding to the administrative areas of Central and North Denmark Regions), covering approximately 1.4 million adults. The study included patients with CKD from 2009 to 2016, and we used data from 1995 to 2008 to assess baseline characteristics and identify participants with prevalent CKD. Health care in Denmark is primarily tax funded, with equal access for every Danish resident (9).

**Data Sources**

We linked data from three databases at the individual level, using a unique personal identifier assigned to every Danish resident (9).

The Danish National Patient Registry contains all primary and secondary diagnoses from inpatient hospitalizations since 1977, and outpatient specialist clinics and emergency room visits since 1995 (10). Diagnoses are recorded according to the International Classification of Diseases Tenth Revision since 1994.

The Laboratory Information System (LABKA) Research Database at Aarhus University contains laboratory test results from every hospital laboratory in the Northern Denmark (11). All laboratory results (excluding some point-of-care tests) from primary, secondary, or tertiary care in Northern Denmark are analyzed and recorded by these hospital laboratories, with information on the type of analysis, date, time, reporting unit, and test result. It was established in 1997 and is considered complete from 2005 (12).

The Civil Registration System contains daily updated data on age, sex, residency, and vital status for all residents of Denmark (13).

The study was reported to the Danish Data Protection Agency (record number 2015–57–00002) by Aarhus University (record number 2016–051–00001/812). According to Danish legislation, approval from an ethics committee or informed consent from patients is not required for registry-based studies.

**Study Population**

We identified all patients with at least one plasma creatinine test in the LABKA database, and retrieved information on these patients from the Danish National Patient Registry, LABKA, and the Civil Registration System. Tests performed in an emergency room or during hospitalization were excluded to allow comparison to international laboratory databases containing only outpatient tests (14,15), and because eGFR is not a validated in AKI, which is more likely during acute illness. For every plasma creatinine, an eGFR was calculated using the CKD Epidemiology Collaboration formula without correction for race (16). Within this source population, we identified patients with first-time CKD using six different algorithms selected from previous studies (Box A1 in Supplemental Appendix) (14,15,17–27). Five of them are on the basis of clinical biochemistry data (laboratory based), while one is on the basis of reported hospital diagnoses consistent with CKD, yielding the following cohorts:

1. The single-test cohort: all patients with at least one reported eGFR <60 ml/min per 1.73 m² (14,17,27).
2. The KDIGO cohort: all patients with at least two reported eGFR <60 ml/min per 1.73 m² ≥90 days apart (6,18–20,27).
3. The KDIGO persistent cohort: all patients with at least two reported eGFR <60 ml/min per 1.73 m² ≥90 days apart with no reported eGFR ≥60 ml/min per 1.73 m² between the two observations (21,22).
4. The KDIGO time-limited cohort: all patients with two records of eGFR <60 ml/min per 1.73 m² between 90 and 365 days apart (15).
5. The KDIGO eGFR/albuminuria cohort: all patients with two reported eGFR <60 ml/min per 1.73 m² or two UACRs >30 mg/g ≥90 days apart (23).
6. The hospital-diagnosed cohort: all patients with a reported primary or secondary diagnosis consistent with CKD, including procedure codes defining dialysis or kidney transplants as recorded in the Danish National Patient Registry during hospital admission, outpatient clinic visit, or emergency room visits (codebook in Supplemental Appendix) (24–26).

The first date of fulfilling the criteria of an algorithm defined the index date (Supplemental Figure 1). Only patients living in the study area for at least a year before the index date were included.

**Statistical Analyses**

We identified age, sex, and Charlson Comorbidity Index conditions recorded in the Danish National Patient Registry at baseline, that is, before the index date (28). Furthermore, we identified the baseline eGFR as the most recent eGFR within the 3 months before or on the index date, and the associated CKD stage by eGFR according to KDIGO G-stages (6). For KDIGO cohorts, we also attained the highest reported eGFR during the CKD defining period of ≥90 days, and the time since a first-reported eGFR <60 ml/min per 1.73 m². For patients in laboratory-based cohorts, we identified the first date of any diagnoses consistent with CKD recorded in hospital, and any procedure code consistent with dialysis or kidney transplant before the index date. Finally, we obtained available information on albuminuria from UACR tests 1 year before or on the index date, using the KDIGO classifications (6): normal or mildly increased (<30 mg/g), moderately increased (30–300 mg/g), or severely increased (>300 mg/g).

For each cohort, we estimated an incidence rate (IR) of CKD with the estimated person time being the sum of adults living in the study area, according to quarterly census numbers from Statistics Denmark (29). The prevalence was estimated as the number of surviving patients with CKD diagnosed by January 1, 2009 per 100,000 adults living in Central and Northern Denmark regions during the last quarter of 2008 (30).

Patients were followed from the index date until death, emigration from the study area, or January 1, 2017. We identified the date of the first-recorded hospital diagnosis
of CKD, and the date of first maintenance dialysis defined as two or more recorded dialysis procedures ≥90 days apart. On the basis of these, we estimated the 1-year cumulative incidence (risk) of death, of receiving a hospital-recorded CKD diagnosis, and of receiving maintenance dialysis (considering competing risk of death). We plotted risk curves for death and maintenance dialysis.

Sensitivity Analyses

To address potential under-reporting of the CKD resulting from the fact that not all adults in Northern Denmark had an eGFR test, we analyzed the incidence and prevalence using only persons with at least one eGFR test as denominator. Also, to estimate the potential effects of excluding test results obtained in emergency rooms or as inpatients, we included these tests in a repeat analysis of the five laboratory-based algorithms (see Box A2 in Supplemental Appendix). Finally, we repeated the analyses of risk of death and maintenance dialysis for patients with CKD G4–G5, after applying the eGFR-based algorithms with an eGFR cutoff of 30 ml/min per 1.73 m². SAS version 9.4 (Cary, NC) were used for all analyses.

Results

We identified 103,435 patients during 2009–2016 with first-time CKD in the single-test cohort, compared with 84,688 patients in the KDIGO cohort, 68,994 patients in the KDIGO persistent cohort, 75,031 patients in the KDIGO time-limited cohort, 100,957 patients in the KDIGO eGFR/albuminuria cohort, and 27,947 in the hospital-diagnosed cohort (Supplemental Figure 2, Table 1). The median age was
between 71 years (interquartile range [IQR], 64–79) and 76 years (IQR, 69–82) across all cohorts, and approximately half the patients were male. The proportion of patients with hospital-recorded comorbidities before the index date differed across the cohorts, with the most common being cerebrovascular disease, chronic pulmonary disease, diabetes, and nonmetastatic solid tumors (Table 1).

The prevalence of CKD varied almost two-fold from 4637 per 100,000 population when using laboratory-based algorithms, and was markedly lower using the hospital-diagnosed algorithm (775 per 100,000 population) (Table 2). Similarly, the annual IR of CKD was highest when defining CKD by a single-test of eGFR (IR, 890 per 100,000 person-years; 95% confidence interval, 885 to 895) and lowest when defining CKD by hospital-diagnoses (IR, 240 per 100,000 person-years; 95% confidence interval, 238 to 243) (Table 2). In a sensitivity analysis including only persons with at least one eGFR test in the denominator, the estimated CKD prevalence and incidences were 1.4-fold and 1.3-fold higher (Supplemental Table 1).

The baseline eGFR levels were similar among the five laboratory-based cohorts with median levels from 54 ml/min per 1.73 m² (in KDIGO persistent) to 58 ml/min per 1.73 m² (in single-test and KDIGO eGFR/albuminuria), and correspondingly, the distribution of patients among CKD stages was comparable in the laboratory-based cohorts. In contrast, baseline eGFR was lower and the distribution among CKD stages very different in patients in the hospital-diagnosed cohort (Figure 1, Table 3). Of note, more than half of patients in the KDIGO and the KDIGO eGFR/albuminuria cohorts had a recorded eGFR >60 ml/min per 1.73 m² during the CKD defining period of ≥90 days (Table 3).

The time from the first-reported eGFR <60 ml/min per 1.73 m² to the index date differed across the cohorts, ranging from 0 months in the single-test cohort (by definition) to 70 months in the hospital-diagnosed cohort (Figure 2, Table 3).

Only a minority of patients had any albuminuria test recorded within 1 year before the index date. Among those, the highest levels of albuminuria were observed in the KDIGO eGFR/albuminuria and hospital-diagnosed cohorts (Table 3).

At 1 year of follow-up, between 7% and 9% of patients across the laboratory-based cohorts died (Table 2), whereas the mortality difference was greater after 8 years, ranging from 39% to 48% (Figure 3). Mortality was much higher in patients with hospital-diagnosed CKD, especially during the first year of follow-up (Figure 3, Table 2). When adjusting for differences in age, the mortality across the laboratory-based cohorts was similar during the entire follow-up (data not shown), whereas the mortality remained higher in the hospital-diagnosed cohort. In each cohort, <0.3% of patients were lost to follow-up during the study period (Supplemental Figure 2).

The 1-year risk of receiving maintenance dialysis was low in all laboratory-based cohorts, and higher in the hospital-diagnosed cohort (Table 2). After 6 years, <1% had...
Table 3. Baseline kidney characteristics of patients with CKD in six cohorts sampled using different algorithms during 2009–2016

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Single Test</th>
<th>Kidney Disease Improving Global Outcomes</th>
<th>Kidney Disease Improving Global Outcomes Persistent</th>
<th>Kidney Disease Improving Global Outcomes Time Limited</th>
<th>Kidney Disease Improving Global Outcomes eGFR/Albuminuria</th>
<th>Hospital Diagnosed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All, N (%)</strong></td>
<td>103,435 (100)</td>
<td>84,688 (100)</td>
<td>68,994 (100)</td>
<td>75,031 (100)</td>
<td>100,957 (100)</td>
<td>27,947 (100)</td>
</tr>
</tbody>
</table>
| **CKD stage on index date, N (%)**
| 1 or 2 | N/A | N/A | N/A | N/A | 15,980 (16) | 7870 (28) |
| 3a | 92,142 (89) | 73,368 (87) | 56,450 (82) | 61,986 (83) | 73,472 (73) | 2899 (10) |
| 3b | 8805 (9) | 9301 (11) | 10,437 (15) | 10,743 (14) | 9310 (9) | 4568 (16) |
| 4 | 1848 (2) | 1652 (2) | 1750 (3) | 1934 (3) | 1654 (2) | 3912 (14) |
| 5 | 640 (0.6) | 367 (0.4) | 357 (0.5) | 368 (0.5) | 376 (0.4) | 1137 (4) |
| Missing | N/A | N/A | N/A | N/A | 174 (0.2) | 7561 (27) |
| **Most recent eGFR 3 mo before or on the index date**, median (IQR) | 56 (52–58) | 63 (58–73) | 57 (53–59) | 59 (55–63) | 67 (59–81) | N/A |
| **Highest eGFR value during CKD defining period, median (IQR)** | 0 | 13 (6–35) | 17 (6–49) | 11 (5–50) | 13 (6–35) | 70 (29–114) |
| **Months since first recorded eGFR <60 ml/min per 1.73 m², median (IQR)** | 573 (0.6) | 580 (0.7) | 532 (0.8) | 615 (0.8) | 589 (0.6) | N/A |
| **Any dialysis in the year preceding CKD, N (%)** | 34 (0.6) | 43 (0.1) | 49 (0.1) | 44 (0.1) | 60 (0.1) | 25 (0.1) |
| **Prevalent CKD diagnosis, N (%)** | 3694 (4) | 4400 (5) | 4211 (6) | 4495 (6) | 5228 (5) | N/A |
| **Months since first CKD diagnosis, median (IQR)** | 36 (4–98) | 24 (5–84) | 21 (4–78) | 18 (4–79) | 29 (5–90) | N/A |
| **Acute or in-patient on the index date, N (%)** | N/A | N/A | N/A | N/A | N/A | 14,144 (51) |
| **UACR categories, N (%)**
| Normal-mild (<30 mg/g) | 12,732 (12) | 11,360 (13) | 9197 (13) | 10,809 (14) | 11,362 (11) | 3386 (12) |
| Moderately increased (30–300 mg/g) | 4465 (4) | 4610 (5) | 3951 (6) | 4588 (6) | 19,304 (19) | 2762 (10) |
| Severely increased (>300 mg/g) | 1131 (1) | 1373 (2) | 1289 (2) | 1462 (2) | 2946 (3) | 1779 (6) |
| Missing | 85,107 (82) | 67,345 (80) | 54,557 (79) | 58,172 (78) | 67,345 (67) | 20,020 (72) |
| U-albumin measured within 365 d before and including index date, N (%) | 14,095 (17) | 17,302 (20) | 14,456 (21) | 16,921 (23) | 33,060 (33) | 8789 (31) |
| U-albumin concentration (mg/L) measured within 365 d before and including index date, median | 14 (7–40) | 15 (7–48) | 16 (7–53) | 16 (7–52) | 36 (13–88) | 37 (10–176) |

IQR, interquartile range; UACR, urine albumin-creatinine ratio.

*On the basis of latest eGFR test during 3 mo before or on index date if available.
commenced maintenance dialysis in any of the laboratory-based cohorts compared with almost 4% in the hospital-diagnosed cohort (Figure 4).

Sensitivity analyses including both in- and outpatient tests produced slightly larger cohorts than the main algorithms, with the associated patient characteristics tabulated in Supplemental Tables 2 and 3. Consequently, the estimated CKD prevalence and incidence were higher (Supplemental Table 4). Of note, the mortality in these sensitivity analysis cohorts varied substantially and was higher than in the main cohorts with 1-year mortalities between 9% and 13% (Supplemental Figure 3, Supplemental Table 4). Finally, when including only patients with CKD stages G4–G5, the different algorithms yielded cohorts with similar risks of death but substantially different risks of maintenance dialysis (Supplemental Figures 4 and 5).

Discussion

Using six different algorithms to identify CKD in medical databases in Northern Denmark, we found large differences in the estimated population prevalence of CKD and in the duration of decreased eGFR at first identification. As expected, a single eGFR test <60 ml/min per 1.73 m² definition of CKD yielded the highest estimated prevalence of CKD. A confirmatory eGFR test after ≥90 days led to a 35% lower estimated CKD prevalence and introduced considerable time (median 13 months; IQR, 6–35) between the first eGFR <60 ml/min per 1.73 m² and the identification of CKD. Baseline characteristics and 1-year mortalities were similar across the five laboratory-based CKD cohorts. In contrast, patients with first-time hospital-diagnosed CKD had a lower median eGFR, longer history of reduced eGFR, greater risk of maintenance dialysis, and a substantially higher mortality.

Previous register-based studies have reported conflicting effects of including a confirmatory eGFR test to define CKD. A Dutch study reported a 24% lower prevalence and a 60% lower incidence when requiring two eGFR tests rather than one (31). Also, a Norwegian study showed that 32% of individuals with a single eGFR <60 ml/min per 1.73 m² had an eGFR ≥60 ml/min per 1.73 m² after ≥90 days (1). In US veterans, CKD prevalence was 27% when on the basis of a single eGFR <60 ml/min per 1.73 m², compared with 10% when requiring two eGFR <60 ml/min per 1.73 m² ≥90 days apart (27). In contrast, studies from the United Kingdom and Sweden reported no substantial differences in the prevalence using algorithms defining CKD by one or more eGFR tests (14,19). These differences may relate to differences in clinical testing practices, laboratory coverage (with some databases limited to general practice, outpatient, or inpatient settings), or by actual difference in the study populations. Nonlaboratory-database studies from Morocco and Belarus showed that 30%–40% of patients with an eGFR <60 ml/min per 1.73 m² at a first test did not reveal a eGFR <60 ml/min per 1.73 m² at systematic follow-up 90 days thereafter (23,32). Many patients in our KDIGO cohort had a test showing eGFR >60 ml/min per 1.73 m² during the CKD defining period of ≥90 days, and thus, a confirmatory test does not per se

Figure 2. | Time since first test of eGFR <60 ml/min per 1.73 m² before index date (month zero) in six cohorts, sampled using different CKD-defining algorithms during 2009–2016 in Northern Denmark.

Figure 3. | Mortality in patients with CKD in six cohorts, sampled using different algorithms during 2009–2016 in Northern Denmark.

Figure 4. | Proportion of patients with CKD receiving maintenance dialysis during <8 years follow-up in six cohorts, sampled using different CKD-defining algorithms during 2009–2016. The intercept of the graph reflects how many were already in maintenance dialysis on the index date.
imply permanently impaired kidney function because eGFR may fluctuate. Importantly, the KDIGO CKD criteria are developed to diagnose patients prospectively during regular clinical examination for CKD, and this may not be the scenario for all persons with eGFR results recorded in laboratory databases.

Interestingly, we found no major differences in 1-year mortality and risk of maintenance dialysis when comparing the different laboratory-based cohorts, whereas the 8-year mortality differed somewhat. This may be due to considerable patient overlap between cohorts concealing potential differences in short-term prognoses. Of note, too strict criteria may misclassify patients with CKD as non-CKD, and may delay the time to the identification of CKD.

Significant differences were identified between the hospital-diagnosed and laboratory-based cohorts. These include a higher 1-year mortality and dialysis risk in the hospital-diagnosed cohort indicating that kidney diseases are not recognized by a relevant diagnosis until significant comorbidity or complications lead to hospital contact.

Also, a substantially higher short-term mortality was observed in the sensitivity cohorts that included inpatient and emergency rooms along with outpatient eGFR tests. This is likely explained by the inclusion of patients with AKI and higher short-term mortality in the cohorts. Thus, we find it appropriate to exclude eGFR results from such visits when identifying CKD.

Although our study was on the basis of routine plasma creatinine tests in a well-defined population with complete follow-up, it has some limitations. Data were not collected as part of a systematic screening and follow-up of the Danish population, and we may underestimate the true incidence and prevalence of CKD when using the entire adult population in the study area as denominator. We addressed this in a sensitivity analysis including only those with a recorded eGFR tests in the denominator, suggesting 1.4-fold higher prevalence and 1.3-fold higher incidence, although these likely overestimate the CKD occurrence in the general population. We did not have information on race for use in the CKD Epidemiology Collaboration formula; however, most Danes (approximately 90%) have a Danish ancestor (30), and thus the potential bias is likely small. When including albuminuria in the algorithm (KDIGO eGFR/albuminuria), the prevalence of CKD was only slightly higher, whereas the incidence of CKD was substantially higher than with the KDIGO algorithm. This is probably due to limited UACR testing before 2009, leading prevalent albuminuria to be classified as incident albuminuria. Furthermore, the higher proportion of diabetes in the KDIGO eGFR/albuminuria cohort than other laboratory-based cohorts may relate to patients with diabetes having more regular albuminuria tests, causing the proportion of patients with diabetes to be higher. Finally, the possibility of dying before a second test may lead to an underestimation of CKD-associated mortality when defining CKD using two eGFR tests. Inker et al. (33) addressed this issue by defining patients who died before a second eGFR test as having CKD; however, this approach most likely overestimates both the prevalence and mortality of CKD.

In conclusion, different algorithms for CKD identification by laboratory parameters greatly affected the size and composition of the CKD cohorts and the estimated prevalence and incidence of CKD, but not the 1-year dialysis and mortality risks, if inpatient and emergency room eGFR tests were excluded. In contrast, an algorithm on the basis of first-time hospital diagnoses of CKD produced a considerably smaller cohort characterized by lower eGFR, longer duration of kidney disease, and higher mortality. Although future studies should clarify if these findings can be generalized to different populations and laboratory testing practices, they point to the importance of recognizing the consequences of using specific algorithms to identify CKD from medical databases.

**Disclosures**

C. F. Christiansen, H. Birn, R. W. Thomsen, S. V. Vestergaard, and U. Heide-Jørgensen report employment with Aarhus University and Aarhus University Hospital. The Department of Clinical Epidemiology, The Department of Biomedicine, and the Department of Renal Medicine at Aarhus University Hospital are involved in studies with funding from various companies as research grants to (and administered by) Aarhus University. None of these are related to this study. H. Birn also reports consultancy agreements with AstraZeneca, Galapagos, and Vifor Pharma; reports receiving research funding from GlaxoSmithKline (GSK) and Vifor Pharma; and reports receiving honoraria from Alexion, AstraZeneca, MSD, Novartis Healthcare, NOVO Nordisk, and Otsuka Pharmaceuticals.

**Funding**

The study was supported by Independent Research Fund Denmark grant 0134-00407B, The Danish Kidney Association, The Danish Society of Nephrology, The Oticon Foundation, and The Hede Nielsen Family Foundation.

**Acknowledgments**

Part of the study was presented at the International Conference on Pharmacoepidemiology and Therapeutic Risk Management 2020.

**Supplemental Material**

This article contains the following supplemental material online at http://cjasn.asnjournals.org/lookup/suppl/doi:10.2215/CJN.15691020/-/DCSupplemental.

- Codebooks.
  - Box A1. Main algorithms.
  - Box A2. Sensitivity algorithms.

- Supplemental Table 1. Prevalence and incidence of CKD when including every person in the background population (white columns) or only persons recorded with available plasma creatinine tests (grey columns) in the denominator.

- Supplemental Table 2. Baseline demographic characteristics and comorbidity before index date of patients with CKD in five sensitivity cohorts sampled using different algorithms during 2009–2016.

- Supplemental Table 3. Baseline kidney characteristics of patients with CKD in five sensitivity cohorts sampled using different algorithms during 2009–2016.


- Supplemental Figure 1. Inclusion of patients to the five laboratory-based cohorts in a hypothetical patient with repeated laboratory tests over time.
Supplemental Figure 2. Flow chart of cohort sampling for six cohorts of patients with CKD in Denmark during 2009–2016.

Supplemental Figure 3. Mortality in patients with CKD in five sensitivity cohorts sampled using different algorithms during 2009–2016 (see Box 2A).

Supplemental Figure 4. Risk of death in patients with CKD stage G4–G5 in four cohorts sampled using different algorithms on the basis of eGFR only during 2009–2016.

Supplemental Figure 5. Risk of maintenance dialysis in patients with CKD stage G4–G5 in four cohorts sampled using different algorithms on the basis of eGFR only during 2009–2016.

References
26. StatisticsDenmark: Population at the first day of the quarter. Available at: www.statbank.dk/F0LK1C. Accessed February 3, 2021


*Received:* October 1, 2020 *Accepted:* January 18, 2021

Published online ahead of print. Publication date available at www.cjasn.org.