# Defining measures of kidney function in observational studies using routine health care data

Carrero, Juan Jesus; Fu, Edouard L.; Vestergaard, Søren V.; Jensen, Simon Kok; Gasparini, Alessandro; Mahalingasivam, Viyaasan; Bell, Samira; Birn, Henrik; Heide-Jørgensen, Uffe; Clase, Catherine M.

*Published in:*
Kidney International

*DOI:*
10.1016/j.kint.2022.09.020



*Publication date:*
2022

# Defining measures of kidney function in observational studies using routine health care data: methodological and reporting considerations

OPEN

Juan Jesus Carrero[1,24], Edouard L. Fu[1,2,3,24], Søren V. Vestergaard[4,5], Simon Kok Jensen[4,5], Alessandro Gasparini[1], Viyaasan Mahalingasivam[6], Samira Bell[7], Henrik Birn[5,8,9], Uffe Heide-Jørgensen[4,5], Catherine M. Clase[10,11], Faye Cleary[6], Josef Coresh[12], Friedo W. Dekker[3], Ron T. Gansevoort[13], Brenda R. Hemmelgarn[14], Kitty J. Jager[15,16], Tazeen H. Jafar[17], Csaba P. Kovesdy[18], Manish M. Sood[19], Bénédicte Stengel[20], Christian F. Christiansen[4,5], Masao Iwagami[6,21,25] and Dorothea Nitsch[6,22,23,25]

[1]*Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Solna, Sweden;* [2]*Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA;* [3]*Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands;* [4]*Department of Clinical Epidemiology, Aarhus University Hospital, Aarhus, Denmark;* [5]*Department of Clinical Medicine, Aarhus University, Aarhus, Denmark;* [6]*Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK;* [7]*Division of Population Health and Genomics, University of Dundee, Dundee, UK;* [8]*Department of Biomedicine, Aarhus University, Aarhus, Denmark;* [9]*Department of Renal Medicine, Aarhus University Hospital, Aarhus, Denmark;* [10]*Department of Medicine, McMaster University, Hamilton, Ontario, Canada;* [11]*Department of Health Research and Methodology, McMaster University, Hamilton, Ontario, Canada;* [12]*Department of Epidemiology, Johns Hopkins University, Baltimore, Maryland, USA;* [13]*Department of Nephrology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands;* [14]*Department of Medicine, University of Alberta, Edmonton, Alberta, Canada;* [15]*ERA Registry, Amsterdam UMC location University of Amsterdam, Medical Informatics, Meibergdreef, Amsterdam, Netherlands;* [16]*Amsterdam Public Health Research Institute, Quality of Care, Amsterdam, the Netherlands;* [17]*Program in Health Services and Systems Research, Duke-NUS Medical School, Singapore;* [18]*Division of Nephrology, Department of Medicine, University of Tennessee Health Science Center, Memphis, Tennessee, USA;* [19]*Department of Medicine, the Ottawa Hospital Research Institute, University of Ottawa, Ottawa, ON, Canada;* [20]*CESP (Center for Research in Epidemiology and Population Health), Clinical Epidemiology Team, University Paris-Saclay, University Versailles-Saint Quentin, Inserm U1018, Villejuif, France;* [21]*Department of Health Services Research, University of Tsukuba, Ibaraki, Japan;* [22]*Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand; and* [23]*UK Renal Registry, UK Kidney Association, Bristol, UK*

**The availability of electronic health records and access to a large number of routine measurements of serum creatinine and urinary albumin enhance the possibilities for epidemiologic research in kidney disease. However, the frequency of health care use and laboratory testing is determined by health status and indication, imposing certain challenges when identifying patients with kidney injury or disease, when using markers of kidney function as covariates, or when evaluating kidney outcomes. Depending on the specific research question, this may influence the interpretation, generalizability, and/or validity of study results. This review illustrates the heterogeneity of working definitions of kidney disease in the scientific literature and discusses advantages and limitations of the most commonly used approaches using 3 examples. We summarize ways to identify and overcome possible biases and conclude by proposing a framework for reporting definitions of exposures and outcomes in studies of kidney disease using routinely collected health care data.**

*Kidney International* (2022) ■, ■–■; https://doi.org/10.1016/j.kint.2022.09.020

KEYWORDS: albuminuria; chronic kidney disease (CKD); creatinine; epidemiology; estimated glomerular filtration rate (eGFR); routinely collected health care data



Routinely collected health care data from registries, electronic health records, and claims databases are increasingly used for research purposes. The availability of laboratory-based kidney function markers, such as serum creatinine and albuminuria, in these data sources increases the opportunities for research in kidney disease. Carefully conducted epidemiologic studies are critical to address the burden, incidence, and prevalence of kidney disease, and to identify mechanisms of action, optimal mitigation/treatment strategies, and gaps in health care processes that collectively have the potential to improve care and, ultimately, outcomes.

**Correspondence:** *Juan Jesus Carrero, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Nobels väg 12A, Box 281, 171 77 Stockholm, Sweden. E-mail: juan.jesus.carrero@ki.se; and Dorothea Nitsch, Department of Non-Communicable Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK. E-mail: Dorothea.Nitsch@lshtm.ac.uk*

[24]JJC and ELF shared first-author position.

[25]MI and DN shared last-author position.

**Received 10 November 2021; revised 31 August 2022; accepted 9 September 2022**

Using routinely collected health care data poses specific challenges when defining chronic kidney disease (CKD),[1] acute kidney injury (AKI),[2] the recently proposed entity, acute kidney disease (AKI that is still evolving),[3] and CKD progression. There are ongoing initiatives to harmonize efforts for establishing diagnoses of CKD, AKI, and acute kidney disease,[4] identify reproducible and valid end points in clinical trials,[5–7] and define outcomes that are important to people living with kidney disease.[8] However, concurrent efforts to harmonize definitions in epidemiologic studies that rely on routine clinical data have been lacking.

The aim of this review is to highlight potential challenges of working definitions of measurements of kidney health in studies of routine care. We start by discussing general issues when working with routine care data: routine care data sources may be fragmented and capture sicker patients. We then discuss pros and cons of the most commonly used definitions and suggest ways to identify and overcome potential biases introduced by using these definitions (Table 1).[9–14] Throughout the article, we use 3 exemplar research questions as illustration. We specifically focus on the following causal questions, for which biases (confounding, selection bias, and measurement bias) are well defined:

(i) What is the causal effect of receiving a CKD diagnosis on mortality risk in older patients with 2 estimated glomerular filtration rate (eGFR) values <60 ml/min per 1.73 $m^2$ >3 months apart?
(ii) Among people with CKD, what is the causal effect of initiating sodium-glucose cotransporter-2 inhibitors versus dipeptidyl peptidase 4 inhibitors on the risk of CKD progression, heart failure admissions, and all-cause mortality?
(iii) After AKI, what is the causal effect of stopping versus continuing renin-angiotensin system inhibitors (RASi) on the risk of recurrent AKI?

We will not cover cohort studies with prospective recruitment and follow-up: considerations for this research design are different from those when using routinely collected data. We conclude by proposing a framework for consensus efforts on reporting definitions of exposures and outcomes in observational studies addressing kidney disease using routinely collected data.

## General issues when working with routine care data

***Data fragmentation affects who is captured and followed up.*** In many countries, routinely collected health care data are captured in disjointed software systems, which are not necessarily integrated.[15] For instance, laboratory information may be only captured in a specific clinical setting,[16] such as ambulatory care, or in hospitals, leading to fragmentation of information and follow-up in the data set. Other databases may include patients on enrollment in an insurance plan (e.g., in the United States), or when they become aged 65 years (e.g., Medicare in the United States and Ontario Drug Benefits): when this happens, data before cohort entry are usually not available. Similarly, patients may exit the database when they move to another general practitioner, or when switching from insurance, not contributing further to the database.[17] Commonly used health care databases in kidney research are summarized in Supplementary Table S1 and Figure 1.[18,19]

The completeness of the data capture may influence the interpretation, generalizability, and internal validity of study results.[20] For instance, when studying the causal effect of initiating sodium-glucose cotransporter-2 inhibitors versus dipeptidyl peptidase 4 inhibitors on CKD progression in claims data sources (question 2), unavailability of data before enrollment may lead to misclassification of conditions, and this may bias effect estimates (information bias).[21] Furthermore, selection bias due to informative censoring[22] will occur when using data sources in which patients with advanced kidney disease get referred from primary care to secondary care and are not followed up thereafter, because those developing advanced kidney disease will drop out from the database.

***Sicker patients have more tests on file than healthy patients.*** Routinely collected health care data do not capture a random sample of the population, but a subgroup of patients who interact with the health care system. As an example, the Stockholm CREAtinine Measurements (SCREAM) cohort in Sweden described that during 2006 to 2011 roughly 67% of the Stockholm population underwent creatinine testing at least once.[23] This nongeneralizability is unlikely to be important for studies that focus on drug effectiveness and safety (such as questions 2 and 3), because the target population for such studies is usually the population eligible for receiving these drugs (i.e., those interacting with the health care system), and not the complete population. However, if the interest would lie in the estimation of the CKD prevalence in Stockholm, the investigator would need to account for the fact that healthier individuals are underrepresented in the data set.[24]

The presence and frequency of a certain laboratory measurement reflect aspects of disease (e.g., albuminuria testing in routine care is mainly directed to specific populations [people with diabetes, hypertension, pregnancy, and known CKD]).[25–27] For question 2, it would be useful to include albuminuria as a potential confounder, but the issue of missing data must be addressed. Excluding individuals without these data may have 2 consequences. First, the cohort may not be similar anymore to the target population, because sicker subgroups are oversampled, which may affect generalizability of study findings (e.g., the medication may be more beneficial in the study population because it oversampled patients with macroalbuminuria, for whom the absolute benefit is larger). It is therefore good practice to report how representative the study population is compared with the target population (e.g., by comparing baseline characteristics or incidence of outcomes).[28] Complete case analysis can lead to bias when data are not missing completely at random (Table 1), although several exceptions exist.[29] We caution against the uncritical use of multiple imputation methods using electronic health data as they can worsen bias if the models are misspecified.[29–32]

**Table 1 | Glossary of terminology associated with bias and examples of research questions**

| | | 1. What is the causal effect of receiving a CKD diagnosis on mortality risk in older patients with 2 eGFR values <60 ml/min per 1.73 $m^2$ >3 months apart? | | 2. Among people with CKD, what is the causal effect of initiating SGLT2i vs. DPP4i on the risk of CKD progression, heart failure admissions, and all-cause mortality? | | 3. After AKI, what is the causal effect of stopping vs. continuing RASi on the risk of recurrent AKI? | |
|---|---|---|---|---|---|---|---|
| | | Exposure: receiving a CKD diagnosis vs. not receiving a CKD diagnosis<br>Outcome: mortality<br>Population: people aged ≥65 yr with 2 eGFR values <60 ml/min per 1.73 $m^2$ >3 months apart | | Exposure: SGLT2i vs. DPP4i<br>Outcomes: CKD progression, heart failure admissions, and all-cause mortality<br>Population: people with CKD | | Exposure: stopping vs. continuing RASi<br>Outcome: recurrent AKI<br>Population: people with AKI | |
| **Bias** | **Definition** | **Example of how bias arises** | **Potential solution** | **Example of how bias arises** | **Potential solution** | **Example of how bias arises** | **Potential solution** |
| Selection bias | Bias in the estimated association or effect of an exposure on an outcome that arises from the procedures used to select individuals into the study or the analysis.<br>It arises when conditioning on a common effect. Examples of selection bias include depletion of susceptibles (or survivor bias), prevalent user bias (those who did not tolerate drug use or died are excluded at baseline), informative censoring or loss to follow up, and missing data.<br>Collider bias is a special case of selection bias where the analysis conditions (either by statistical adjustment or restriction of the study population) on a collider (a variable that is affected by 2 other variables [e.g., exposure or outcome or related variables, which then introduce a spurious association]). | 1. Assessing the relationship between CKD diagnosis and mortality in routinely collected health care data implicitly restricts to the subset of people with 2 eGFR measurements. As both eGFR level and health status influence availability of test results in the study, collider bias is introduced (Supplementary Figure S1A).<br>2. When follow-up is started from the second measurement, depletion of susceptibles may lead to selection bias due to 2 colliders (Supplementary Figure S1B). | 1. Conduct study in a setting/population where kidney function is measured in everybody at baseline (e.g., by restricting to a certain subpopulation).<br>2. Selection bias due to depletion of susceptibles may be small when the window between 2 measurements is short and low-risk populations are studied. | 1. Loss to follow-up or dropout that is differential with respect to the exposure leads to selection bias (e.g., this occurs when data are fragmented [only primary care data are available] and more people in the DPP4i arm are lost to follow-up as they transition to specialized care). Differential dropout due to death also biases estimation of kidney function slopes (Supplementary Figure S1C).<br>2. Missing data: excluding individuals without eGFR or UACR measurements introduces selection bias when the missingness depends on the exposure (either directly or indirectly) and health status (Supplementary Figure S1D). | 1. Loss to follow-up or dropout can be handled using inverse probability of censoring weighting or by using joint models, which explicitly model the dropout process and longitudinal outcome simultaneously through shared random effects.<br>2. Provide clarity as to who the population with available measurements (eGFR/UACR measurements) was and do not extrapolate further. Multiple imputation can be attempted but may be misspecified. | Studying recurrent events is susceptible to selection bias when prior treatment influences the risk of AKI[9] (Supplementary Figure S1E). | Adjusting for prior treatment, or restricting to people who have not been treated previously, removes the selection bias. |

(*Continued on following page*)

**Table 1 |** (Continued) **Glossary of terminology associated with bias and examples of research questions**

| | | 1. What is the causal effect of receiving a CKD diagnosis on mortality risk in older patients with 2 eGFR values <60 ml/min per 1.73 $m^2$ >3 months apart? | | 2. Among people with CKD, what is the causal effect of initiating SGLT2i vs. DPP4i on the risk of CKD progression, heart failure admissions, and all-cause mortality? | | 3. After AKI, what is the causal effect of stopping vs. continuing RASi on the risk of recurrent AKI? | |
|---|---|---|---|---|---|---|---|
| | | Exposure: receiving a CKD diagnosis vs. not receiving a CKD diagnosis<br>Outcome: mortality<br>Population: people aged ≥65 yr with 2 eGFR values <60 ml/min per 1.73 $m^2$ >3 months apart | | Exposure: SGLT2i vs. DPP4i<br>Outcomes: CKD progression, heart failure admissions, and all-cause mortality<br>Population: people with CKD | | Exposure: stopping vs. continuing RASi<br>Outcome: recurrent AKI<br>Population: people with AKI | |
| **Bias** | **Definition** | **Example of how bias arises** | **Potential solution** | **Example of how bias arises** | **Potential solution** | **Example of how bias arises** | **Potential solution** |
| Information bias | Bias in an estimate arising from measurement errors or misclassification (the erroneous classification of an individual, a value, or an attribute into a category other than that to which it should be assigned). | Immortal time bias: immortal time is introduced if follow-up is started at the moment eGFR decreases to <60 ml/min per 1.73 $m^2$, but patients who receive a CKD diagnosis later during follow-up are classified into the CKD diagnosis group. Immortal time bias can be considered a form of misclassification because unexposed person-time is incorrectly considered exposed person-time. | Causal study designs, such as clone-censor-weight or sequential trials, appropriately align the start of follow-up with the start of exposure and mitigate immortal time bias. | 1. More kidney function measurements are taken during follow-up in the DPP4i arm, leading to differential measurement error in the outcome. CKD progression is therefore more likely to be picked up in the DPP4i arm, biasing the effect estimates.<br><br>2. Admissions of patients with advanced CKD who have volume overload but normal cardiac function on echocardiography are miscoded as admissions with heart failure.<br><br>3. Classifying transient decreases in GFR as CKD progression leads to bias in the estimation of the incidence rate of the outcome and of the absolute risk differences. | 1. To detect differential outcome ascertainment bias, the number of kidney function measurements can be compared between exposure groups. More kidney function measurements in 1 exposure group than the other point toward differential outcome ascertainment bias. Differential outcome ascertainment bias is unlikely to occur for hard end points that do not depend on testing, such as kidney replacement therapy.<br><br>2. Perform validation study of heart failure codes for population with CKD to quantify whether this bias is common or rare. The obtained sensitivity and specificity can be used in quantitative bias analyses to provide adjusted effect estimates.<br><br>3. Identifying sustained declines in eGFR by using a linear mixed model helps to appropriately classify transient decreases in GFR as nonevent. | 1. Using diagnosis codes to ascertain AKI may miss many AKI events (high specificity and low sensitivity), [10–13] leading to an underestimate of the incidence rate and bias in absolute risk differences<br><br>2. Differential outcome ascertainment may occur if more creatinine measurements are performed during follow-up in 1 exposure group and hospitalizations without a baseline creatinine measurement are not considered for AKI events. | Clearly report the definition used. Use different definitions to assess their influence on point estimates. To detect differential outcome ascertainment bias, the number of kidney function measurements can be compared between exposure groups. |
| Confounding bias | Bias of the estimated effect of an exposure on an outcome because of the presence of common causes of the exposure and the outcome. | Many risk factors for receiving a CKD diagnosis are also risk factors for all-cause mortality. For instance, physicians may give a coded CKD diagnosis to sicker patients who have more comorbidities and a lower eGFR. | Measuring and appropriately adjusting for all confounders. Alternatively, quasi-experimental designs, such as regression discontinuity, can be used to study the effect of receiving a CKD diagnosis on outcomes by using a "threshold" (i.e., the probability to receive | SGLT2i are more likely to be prescribed in people with CKD (study population) but with prescriptions depending on kidney function itself, and also in people with atherosclerotic cardiovascular disease, or heart failure.<br>Corresponding diagnosis codes (which often have high specificity but low sensitivity) may lead to residual confounding. | Measure and adjust for all confounders. Whenever available, adjust for measurements of kidney function, such as eGFR and UACR, and metrics of heart failure/volume overload at baseline (e.g., LVEF and NT-proBNP). Be aware of fragmentation of data.<br>Negative or positive control outcomes can be used to detect and adjust for residual confounding.<br>Quantitative bias analysis can be used to assess the influence of residual confounding on effect estimates. | The severity of AKI may be an important confounder because it will influence the decision of whether to stop RASi and is associated with the likelihood of having a repeated AKI. | Adjust for the severity of AKI, taking into account the magnitude of creatinine elevations as well as whether kidney replacement therapy was needed. |

(*Continued on following page*)

**Table 1 |** (Continued) **Glossary of terminology associated with bias and examples of research questions**

| Bias | Definition | 1. What is the causal effect of receiving a CKD diagnosis on mortality risk in older patients with 2 eGFR values <60 ml/min per 1.73 $m^2$ >3 months apart? | | 2. Among people with CKD, what is the causal effect of initiating SGLT2i vs. DPP4i on the risk of CKD progression, heart failure admissions, and all-cause mortality? | | 3. After AKI, what is the causal effect of stopping vs. continuing RASi on the risk of recurrent AKI? | |
|---|---|---|---|---|---|---|---|
| | | Exposure: receiving a CKD diagnosis vs. not receiving a CKD diagnosis<br>Outcome: mortality<br>Population: people aged ≥65 yr with 2 eGFR values <60 ml/min per 1.73 $m^2$ >3 months apart | | Exposure: SGLT2i vs. DPP4i<br>Outcomes: CKD progression, heart failure admissions, and all-cause mortality<br>Population: people with CKD | | Exposure: stopping vs. continuing RASi<br>Outcome: recurrent AKI<br>Population: people with AKI | |
| | | Example of how bias arises | Potential solution | Example of how bias arises | Potential solution | Example of how bias arises | Potential solution |
| | | a CKD diagnosis is much higher <60 ml/min per 1.73 $m^2$ than >60 ml/min per 1.73 $m^2$, but people closely around the threshold have a similar prognosis). | | | | | |

AKI, acute kidney injury; CKD, chronic kidney disease; DPP4i, dipeptidyl peptidase 4 inhibitors; eGFR, estimated glomerular filtration rate; GFR, glomerular filtration rate; LVEF, left ventricular ejection fraction; NT-proBNP, N-terminal prohormone of brain natriuretic peptide; RASi, renin-angiotensin system inhibitors; SGLT2i, sodium-glucose cotransporter-2 inhibitors; UACR, urinary albumin–to–creatinine ratio.
Definitions were sourced from *Last's Dictionary of Epidemiology*[14] and adapted to the current context.

Laboratory testing may also be influenced by external factors, such as financial incentivization. For instance, there was a notable increase in serum creatinine testing among patients with diabetes attending primary care in the United Kingdom following the implementation of the Quality and Outcomes Framework in 2004.[33]

## Considerations when using CKD as exposure or population

Various algorithms have been used to identify persons with CKD in health care databases[34]: Table 2 describes those most commonly used, along with identified merits and caveats. Figure 2 graphically shows an example of how different algorithms may identify the same patient at different points during the disease course. This means that for research question 2, different CKD populations will be identified, depending on the definition used, which affects generalizability and interpretation of study results.

***Diagnostic coding of CKD.*** In settings without laboratory data, diagnosis codes (e.g., *International Classification of Diseases, Ninth Revision* [*ICD-9*] or *International Classification of Disease, Tenth Revision* [*ICD-10*]) are commonly used to identify patients with CKD.[34] Diagnostic codes have high specificity for CKD, and can detect patients with structural abnormalities not recognized by laboratory-based algorithms.[33–35] However, relying on recorded clinical diagnoses of CKD often fails to identify a large proportion of patients with CKD due to limited awareness of kidney disease, meaning a low sensitivity.[36,37] The consequences of using diagnostic codes to identify patients with CKD also depends on coding practices: increasing awareness resulting from system changes, such as automatic eGFR, can lead to changes in the completeness of data over time.[38] In studies with cohort identification periods spanning many years, underlying morbidity or severity of diagnosed CKD in selected patients may vary over time.[39–42] For questions 1 and 2, studies should therefore take account of calendar year and health provider (e.g., different general practitioners in the United Kingdom) to address temporal and health provider variation in CKD identification, which is likely nonrandom, and potentially associated with health outcomes. We suggest using the term "diagnosed CKD" when detection is limited to *International Classification of Diseases* codes.

For question 1, "What is the causal effect of receiving a CKD diagnosis on mortality risk in older patients with 2 eGFR values <60 ml/min per 1.73 $m^2$ >3 months apart?," receiving a CKD diagnosis is the exposure, but not the population. Those who have biochemical evidence for a CKD diagnosis but have no formal diagnosis on file are the comparison group.

***Laboratory variables and equations to estimate glomerular filtration rate.*** The laboratory assay used for quantifying serum or plasma creatinine, and its traceability to the isotope-dilution mass spectrometry international standard, as well as the equation used for estimating glomerular filtration rate should be clearly reported in research.[43] Researchers need to be aware that eGFR may not reflect true kidney function; and

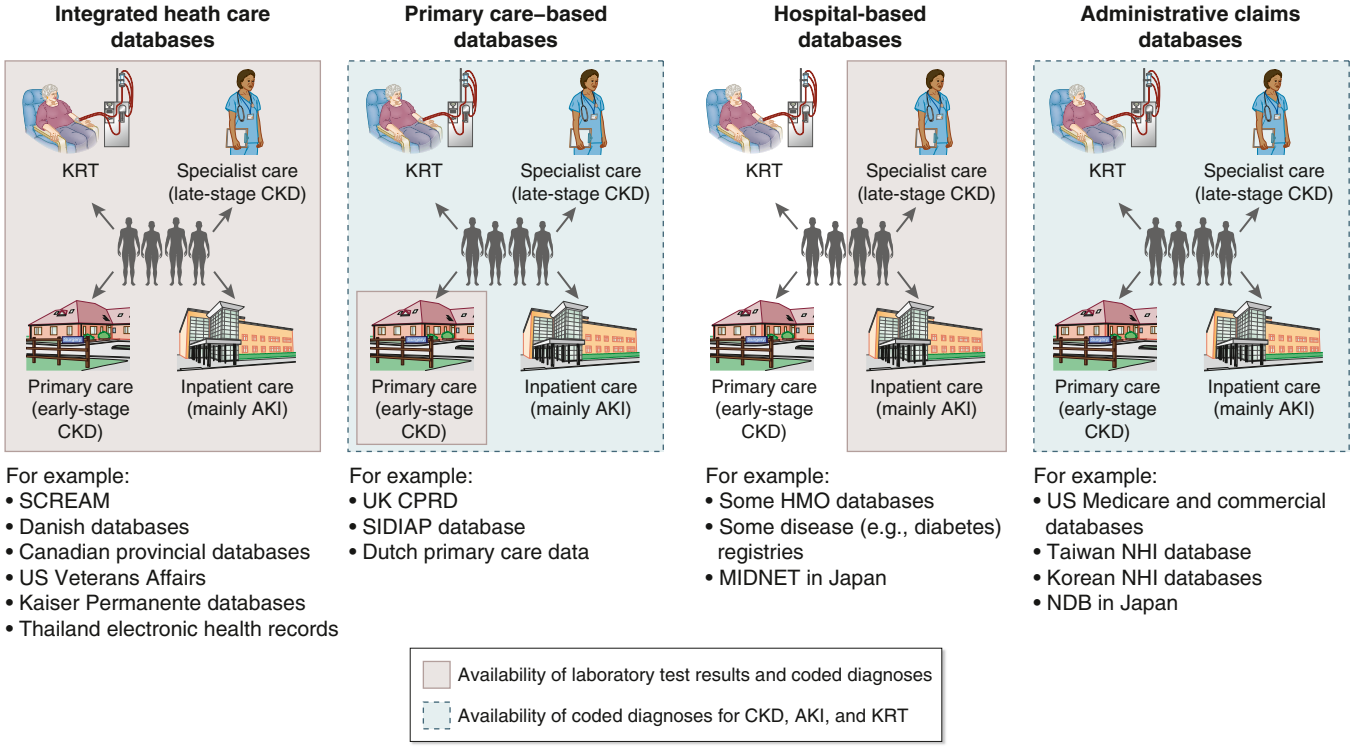


**Figure 1 | Overview of different routinely collected health care databases used in kidney disease research, illustrating data fragmentation.** AKI, acute kidney injury; CKD, chronic kidney disease; CPRD, Clinical Practice Research Datalink; HMO, health maintenance organization; KRT, kidney replacement therapy; MIDNET, Medical Information Database NETwork; NDB, National Database of Health Insurance Claims and Specific Health Checkups of Japan; NHI, National Health Insurance; SCREAM, Stockholm CREAtinine Measurements; SIDIAP, Information System for Research in Primary Care in Catalonia, Spain.

depending on the study question, this can lead to bias. An additional issue, if the data are available, is the type of health care encounter in which the test took place (i.e., outpatient vs. inpatient creatinine test). We note that eGFR equations have not been validated when kidney function is not stable.

***Consequences of using the chronicity criterion.*** Kidney Disease: Improving Global Outcomes (KDIGO) developed a consensus definition for diagnosing CKD in clinical practice.[1] This definition is based on the presence of reduced eGFR or albuminuria for at least 3 months or structural abnormalities of the kidneys, and requires repeated testing, if the first screening result of eGFR or albuminuria is abnormal. Ensuring chronicity is essential in establishing a CKD diagnosis and has become routine worldwide. This approach has also frequently been applied in observational studies evaluating the incidence, prevalence, risk factors, and outcomes of CKD.[44–47] However, requiring 2 consecutive eGFR measurements 3 months apart in routinely collected data can lead to a selective population, because it requires that the patient is sick enough to seek health care twice or be recalled for a confirmatory test, which may vary between clinicians and based on patient characteristics. Furthermore, if the time for the diagnosis of CKD is defined by the second low eGFR beyond 3 months, the identification with CKD will be delayed at least 3 months. In a recent study, using the definition for CKD of 2 eGFR measurements of $<60$ ml/min per 1.73 $m^2$ at least 90 days apart (with no upper limit) resulted in a "delay," with more than half of patients being recognized as having CKD $>1$ year after the first low eGFR (median, 13 months [interquartile range, 6–35 months]).[37]

When answering research question 1, care should be taken to appropriately align the start of follow-up with the start of the exposure to prevent immortal time bias or depletion of susceptibles bias.[48] If follow-up is started at the time eGFR decreases below 60 ml/min per 1.73 $m^2$, but patients receive a diagnosis of CKD only later during follow-up, immortal time will be introduced.[49–51] Patients in the CKD diagnosis group cannot die during the period between eGFR $<60$ ml/min per 1.73 $m^2$ and the CKD diagnosis. After all, they would have been assigned to the "no CKD diagnosis" group if they had died during this period. This gives an artificial survival advantage to the CKD diagnosis group. If receiving a CKD diagnosis truly has a causal effect (either beneficial or harmful) on mortality, and follow-up is started some period *after* patients received a CKD diagnosis, "depletion of susceptibles" bias is introduced,[52] which is a form of selection or

**Table 2 | Advantages and disadvantages of different definitions of CKD used in previous studies based on routinely collected data**

| Definition | Advantages | Disadvantages |
|---|---|---|
| Diagnosis codes | - High specificity, because clinically verified<br>- Usually available in data sources without laboratory measurements (e.g., claims databases)<br>- May pick up structural changes that are not picked up by eGFR and/or albuminuria definitions | - Low sensitivity<br>- Considerable delay in identification<br>- Sensitive to changes in testing and coding practices<br>- Misclassification influenced by coding practices and purpose (e.g., reimbursement, pay for performance, and documentation in routine practice) |
| Single eGFR <60 ml/min per 1.73 $m^2$<br>Single UACR >30 mg/g | - High sensitivity<br>- Minimal delay in identification | - Sensitive to changes in testing practices<br>- Loss of information associated with dichotomizing the outcome by a certain threshold<br>- Lacks confirmation of chronicity<br>- May identify AKI or AKD instead of CKD<br>- Testing for albuminuria is less frequent and may vary between specific patient groups; selected patient groups tested for UACR will be overrepresented |
| Two eGFRs <60 ml/min per 1.73 $m^2$ and/or 2 UACRs >30 mg/g at least 90 d apart | - Ensures chronicity<br>- In accordance with guidelines<br>- Acknowledges the criteria of kidney damage | - Delay/missed identification (requires regular testing in study population)<br>- Sensitive to changes in testing practices<br>- May identify patients with 2 episodes of AKI or dehydration. Additional condition "no eGFR >60 ml/min per 1.73 $m^2$ or UACR <30 mg/g during the CKD-defining period of at least 90 d" could minimize the risk of including such patients<br>- A time limit (e.g., no more than 365 d apart) may need to be defined to target well-observed patients with CKD, in return for higher risk of missing patients with infrequent tests<br>- Baseline for follow-up can only start at second measurement, resulting in survivor bias<br>- Testing for albuminuria is less frequent and may vary between specific patient groups; selected patient groups tested for UACR will be overrepresented |

AKD, acute kidney disease; AKI, acute kidney injury; CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate; UACR, urine albumin–to–creatinine ratio.

survivorship bias (Table 1). Analyses will need to use appropriate statistical methods (e.g., time-dependent exposure variables) or comparison groups (e.g., allowing for the same period of "run-in" immortal time for the non-CKD cohort) to align the start of follow-up and start of exposure to prevent immortal time bias and depletion of susceptibles bias.[16,48,53,54]

***Adding albuminuria to eGFR to classify CKD.*** Most observational studies identify CKD cases on the basis of eGFR only.[55] At least a fifth of the populations with CKD remain understudied and uncharacterized because they have CKD category G1 and G2 and require either A2 or greater urine albumin–to–creatinine ratio (UACR) or other signs of renal damage (e.g., structural kidney disease) for identification.[56,57] Recent initiatives have been taken to improve patient identification in routine databases by developing conversion formulas between urinary protein-creatinine ratio or urinary dipstick protein to UACR.[58,59] However, even these tests are not universally performed, and even if such conversion is introduced, the tests are not fully comparable. Notably, urine dipstick analysis, which measures protein, but not creatinine, has generally high false-negative rates, and can also have a high false-positive rate in the general community setting when compared with more quantitative tests.[60] Finally, researchers have used different strategies for classification of CKD (e.g., the least severe, the most severe, the most recent, or the mean or median of eGFR or UACR level during the period used to define CKD).[61,62] Being transparent about and justifying the chosen definition are essential for the reader to understand the study's strengths, limitations, generalizability, and likely reproducibility.

## Defining CKD progression

There is ample heterogeneity in how CKD progression is defined in epidemiologic studies, including both claims-related end points (kidney replacement therapy [KRT][53] or death attributed to CKD[63]), time to laboratory-based percentages of eGFR decline relative to baseline (typically 30%, 40%, 50%, or 57%),[64] time to doubling of serum creatinine,[65] eGFR values below a certain threshold (e.g., incident <60 or <15 ml/min per 1.73 $m^2$),[66] diagnostic coding for CKD,[67] longitudinal eGFR decline, and combinations of these in a composite outcome. Table 3 lists some of the methods used to define CKD progression and discusses pros and cons. The same challenges that apply to CKD ascertainment also apply herein. Because of space limitations, we will not discuss definitions of albuminuria progression, which can be ascertained by transition to a different "A" category or changes in continuous UACR over time. As explained earlier, the capacity to detect these outcomes depends on the type of testing

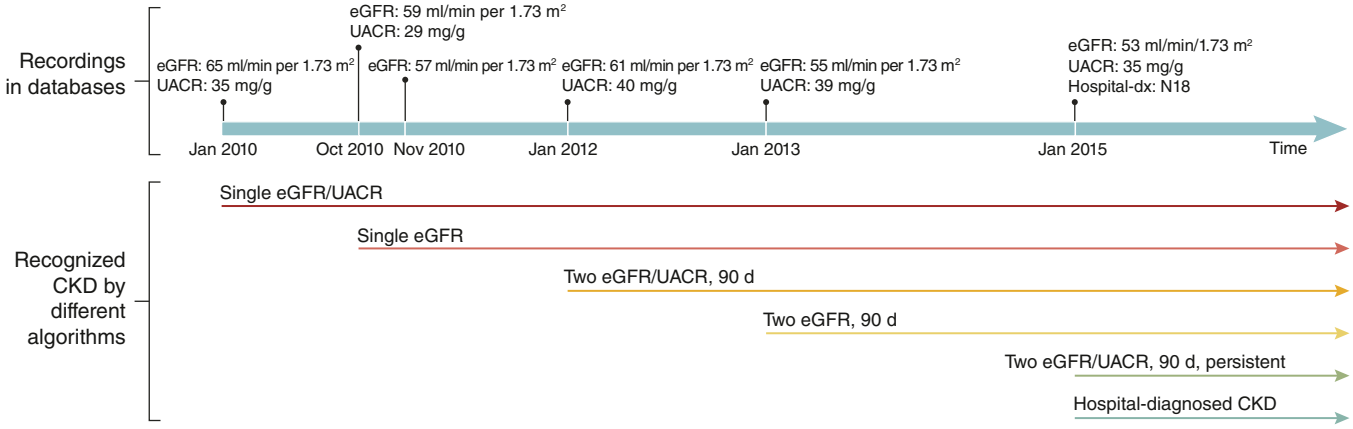


**Figure 2 | The algorithm used to identify chronic kidney disease (CKD) influences when patients are included in the study.** This is an example of a patient with recorded estimated glomerular filtration rate (eGFR) and urine albumin–to–creatinine ratio (UACR) tests and hospital diagnoses (dx; top), and common algorithms that would recognize the patients as having CKD based on the recordings (below). In this case example, there is a 5-year time gap between identification of CKD by the first and the last defining algorithm, during which the patients need to survive.

(categories of albuminuria by dipstick or continuous albuminuria concentration by UACR) and frequency of testing and is limited by the fact that testing tends to be directed toward people at higher risk.

***Kidney replacement therapy.*** KRT has historically been the preferred outcome for observational studies, because it is assumed that the incidence of KRT is not affected by the frequency of laboratory testing or health care use, and it is clearly an outcome of great importance to patients.[8] Thus, bias due to differential outcome ascertainment is unlikely to occur when using KRT.[68] However, KRT is not the same as eGFR <15 ml/min per 1.73 $m^2$—many, especially older, patients with CKD G5 survive for a prolonged period without requiring dialysis, or choose not to have dialysis, either temporarily or as an enduring decision (which in clinical practice is sometimes termed conservative care). Hence, those who access KRT are a selected group of patients, and, depending on health system funding of KRT, may not represent the population burden of people most severely affected by kidney disease. Ascertainment of KRT episodes in administrative data requires algorithms to identify the date of the first chronic dialysis. Alternatively, data sources in selected countries often have linkages with national KRT registries. Finally, a potential disadvantage is that evaluating risks of KRT may require large sample size and long follow-up for sufficient power in low-risk populations.

***Laboratory-based definition of CKD progression.*** Most studies use composite outcomes that incorporate creatinine- or eGFR-based definitions. As discussed above, the frequency, indication, and location of testing may pose a risk for differential outcome ascertainment. For question 2, this differential outcome ascertainment occurs when there are more creatinine measurements in the dipeptidyl peptidase 4 inhibitor arm than in the sodium-glucose cotransporter-2 inhibitor arm. The current consensus is to consider a 30% to 40% glomerular filtration rate decline as a surrogate end point for kidney failure for clinical trials of CKD progression,[69,70] and these are also often applied in observational studies. Finally, these surrogate outcomes are chosen with the idea of reflecting a clinically important event, but the dichotomization comes at the expense of loss of information and loss of power.[71]

***Sustained declines of eGFR over time.*** Figure 3 illustrates the trajectory of outpatient eGFR measurements over time from selected participants in the SCREAM project. It can be easily observed that in some cases, reaching a certain threshold does not necessarily align with the behavior of the rest of kidney function measurements throughout the patient's journey. Nonrenal determinants of eGFR, intense periods of disease and testing, or even an AKI episode may falsely be identified as a doubling of creatinine or as a 30% eGFR decline from baseline. For instance, the incidence rate of CKD progression may be overestimated in question 2, when only 1 measurement below a certain threshold is required to be considered as the occurrence of an outcome.

A confirmation measurement (i.e., a decline in eGFR that is sustained over time) will improve the positive predictive value of the outcome, at a cost. How the scientific literature addresses this is variable, and often not reported. The considerations around confirmation, mentioned earlier in the indentation patients with CKD, apply herein; requiring the presence of consecutive measurements of a similar magnitude or relative eGFR reduction depends on health care access and testing,[72] and is not possible in the case of death.

In practice, researchers sometimes fit a linear regression line through the eGFR measurements that are available per individual to confirm a sustained decline and ascertain when a certain eGFR threshold is reached.[73,74] However, linear regression cannot be estimated well if only few measurements are available, and often patients with only 1 eGFR measurement during follow-up are excluded.[74] Furthermore, people may drop out owing to KRT or death. A better alternative is to

**Table 3 | Advantages and disadvantages of different definitions of CKD progression when using routinely collected health care data**

| Definition | Advantages | Disadvantages |
|---|---|---|
| Diagnosis codes for more severe CKD stages | - High specificity, because clinically verified<br>- Available in settings without laboratory data | - Low sensitivity<br>- Considerable delay in identification (codes may not be updated regularly to reflect kidney function change)<br>- Changes as coding practices/incentives change<br>- Dependent on physician awareness, likely to be highest in patients who seek care more often<br>- May distort measures of inequality if a particular group is less likely to be diagnosed (e.g., women and ethnic minorities). |
| Initiation of KRT | - Hard end point and of great importance to patients<br>- Strongly related to cost of care<br>- Low likelihood of differential outcome ascertainment | - May not be available without linkage to national registry<br>- Subject to clinical judgement/practice variation<br>- National registry may not capture all acute dialysis starters (typically only KRT rates for 90-d survivors are reported)<br>- Will only capture those who are offered and elect to undergo dialysis/transplantation<br>- Not valid in settings where economic inequalities and absence of funding make KRT unaffordable to many patients<br>- In low-risk populations, too few events, resulting in underpowered study<br>- In view of high competing mortality, less informative for early prevention efforts |
| eGFR <15 ml/min per 1.73 $m^2$ without confirmation | - Better proxy for kidney failure than KRT<br>- Many patients with this level of kidney function will present to health services because of symptoms | - To distinguish new decline from undetected long-standing CKD, this can only be used in a population who undergoes repeated kidney function testing<br>- Depends on who has access to test (setting and funding)<br>- Susceptible to measurement error<br>- May identify AKI instead of CKD<br>- Interpretation in terms of cost implications/health burden can be different from the interpretation of KRT, particularly at older age |
| eGFR <15 ml/min per 1.73 $m^2$ with confirmation | - Better proxy for kidney failure than KRT<br>- Includes conservative care<br>- Applies in LMIC, where KRT may not be available or universally accessible | - To define an incident event, this requires a population that undergoes repeated kidney function testing and depends on who has access to test (setting and funding).<br>- Competing mortality (high risk of death after first eGFR <15ml/min per 1.73 $m^2$).<br>- Interpretation in terms of cost implications/health burden can be different from the interpretation of KRT, particularly at older age |
| Time to % eGFR decline (30%, 40%, 50%, or 57%) without confirmation | - More power and greater relevance for early prevention at higher CKD GFR stages<br>- Larger eGFR declines better surrogate measure for kidney failure | - May identify AKI instead of CKD<br>- Susceptible to measurement error<br>- Some events are transient because of eGFR fluctuations<br>- Loss of information associated with dichotomizing the outcome by a certain threshold |
| Time to % eGFR decline (30%, 40%, 50%, or 57%) with confirmation | - More power and greater relevance for early prevention at higher CKD GFR stages<br>- Larger eGFR declines better proxy for kidney failure<br>- More robust to transient changes in eGFR | - Same as above and also:<br>- Delay in identification or failure to identify in case of death<br>- Immortal time<br>- Informative visit process as timing to next test driven by patient characteristics/comorbidity status |
| Linear interpolation and smoothing of eGFR slopes with linear regression | - Uses all measurements, so less sensitive to AKI or measurement error<br>- Easy to implement<br>- May be accurate when using prospective data with no dropout and at least 3 measurements per person | - Performance likely to be worse than linear mixed models in routinely collected data because of few measurements and dropout<br>- If only few measurements are available, the slope cannot be estimated well and hence the time point of crossing the threshold cannot be precisely determined<br>- Patients with only 1 measurement during follow-up are excluded<br>- Gives biased estimates in the case of dropout due to kidney failure with replacement therapy or death |
| Longitudinal eGFR decline with linear mixed model | - Superior performance to linear regression<br>- Uses all measurements, so less sensitive to AKI or measurement error<br>- Can account for data missing at random<br>- Can include patients with only 1 measurement or few measurements<br>- Fitted model can be used to ascertain when a certain decline threshold was reached (sustained decline) | - Reasons for repeated testing can bias coefficients associated with random effects (severe bias only when all measurements are irregular); explicit modeling assumptions required to address competing mortality and informative censoring in joint models<br>- Assumes linear eGFR decline, but the linearity assumptions can be relaxed by including appropriate transformations of time in the model. |
| Progression of albuminuria | - Part of KDIGO CKD definition<br>- Often assessed in clinical trials<br>- Formulas have been developed to convert urinary PCR or dipstick measurements to ACR | - Substantial bias by reasons for urine testing<br>- High variability of albuminuria introduces substantial measurement error (difficult to interpret small changes at the individual level) |

ACR, albumin-to-creatinine ratio; AKI, acute kidney injury; CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate; GFR, glomerular filtration rate; KDIGO, Kidney Disease: Improving Global Outcome; KRT, kidney replacement therapy; LMIC, low- and middle-income countries; PCR, protein-to-creatinine ratio.

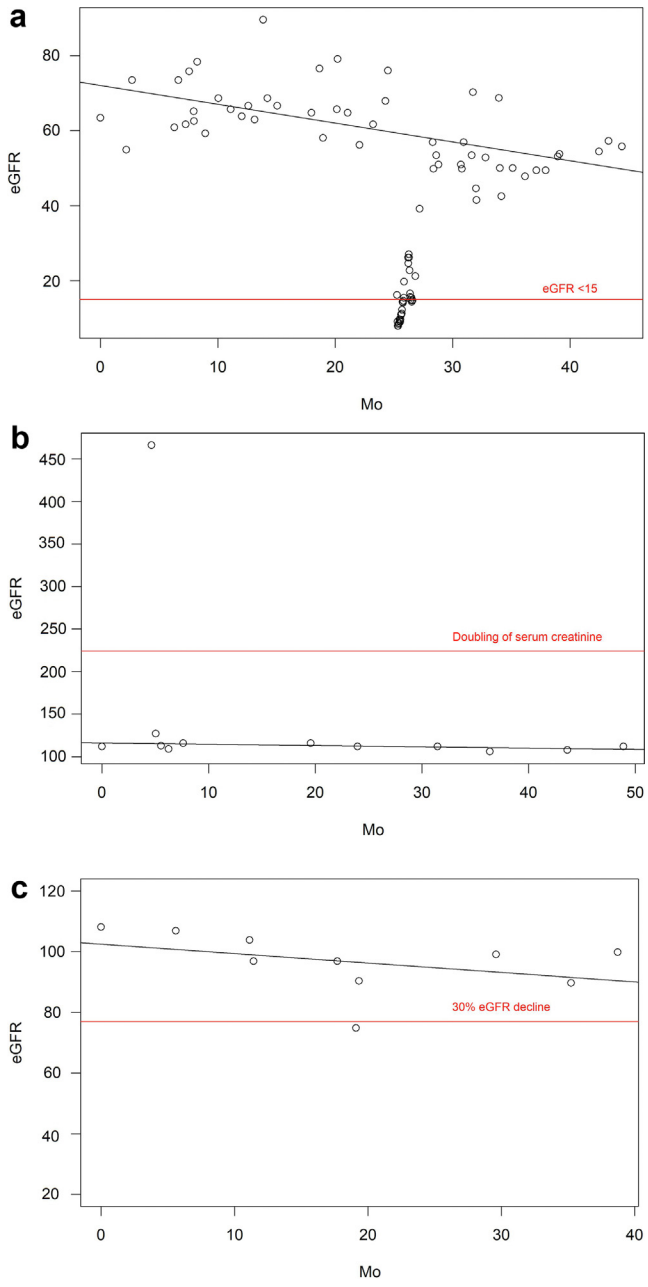


**Figure 3 | Plots of outpatient estimated glomerular filtration rate (eGFR) or creatinine measurements from 3 individuals in the Stockholm Creatinine Measurements database.** (**a**) The red line depicts an eGFR <15 ml/min per 1.73 $m^2$. This individual has many measurements during follow-up. At 25 months, there are many low eGFR measurements, which may represent acute illness or an acute kidney injury (AKI). Identifying simply a decline of >30% from baseline as study outcome would misclassify this AKI as a chronic kidney disease (CKD) progression event, leading to a biased estimate of the incidence of CKD progression in the study population. (**b**) The individual has many creatinine measurements after baseline, with 1 creatinine measurement surpassing the threshold of doubling of serum creatinine (red line). However, this likely does not reflect a "true" doubling of serum creatinine. Note that the y-axis suggests serum creatinine (umol/L). (**c**) The individual has 10 eGFR measurements, with 1 measurement below the threshold of 30% eGFR decline. On the basis of the global information that we have for this patient, it seems a random observation possibly influenced by disease or hydration status.

use a linear mixed model with random intercepts and slopes.[74–76] The use of random effects allows for "borrowing" of information across individuals. These models increase stability for patients for whom only few measurements are available and can include information from individuals with only 1 follow-up eGFR assessment. Furthermore, under certain conditions, mixed models can even handle informative missingness if the predictors of missingness are included as covariates in the model.[77] It is recommended to add the baseline eGFR value to the outcome vector. Linearity assumptions can be easily relaxed by including appropriate transformations of time in the model, such as quadratic terms or splines. The fitted mixed model can then be used to ascertain when a certain decline threshold was reached. In this definition, reaching the interpolated threshold will be sustained over time, and may identify outcomes that theoretically occurred earlier, during a period when laboratory testing was not frequent. One can easily see how interpolation with smoothing of the eGFR slope can also serve to improve the identification of patients with "confirmed" CKD.

### Considerations for choosing AKI definitions

The scientific literature reports varying algorithms to define AKI in health care databases (Table 4).[78] Below, we summarize common features of these definitions that require attention as they can affect the generalizability or validity of study findings. We illustrate this using our third example question: "After AKI, what is the causal effect of stopping versus continuing renin-angiotensin system inhibitors on the risk of recurrent AKI?"[79] For this particular research question, the inception episode of AKI defines the population of interest, the next episode is the outcome of interest, and history of AKI before inception could be used as a covariate to adjust for confounding. Most of the considerations discussed below likely also apply when evaluating the newly defined entity of acute kidney disease,[3,4] but few studies to date have explored acute kidney disease in health care databases.[80]

***Diagnostic coding of AKI.*** Hospital-recorded AKI diagnoses are often included in health care databases and coded with *International Classification of Diseases* codes, and these are often used to identify AKI populations, as an outcome or as a covariate. Although the specificity of the hospital-recorded diagnoses is high (>95%), the coding is incomplete and may only identify a quarter to a third of all AKI episodes identified by changes in serum creatinine,[10–13] even fewer when considering all cases, including those defined by oliguria.[13] The reason for coding (e.g., reimbursement, pay for performance, or documentation in routine practice) may also impact the validity of codes. For our specific example, using AKI diagnosis codes will lead to a selective population of more severe AKI cases, which may impact the generalizability of results: findings may not be necessarily generalized to the complete AKI population, and would also include more severe AKI cases (i.e., stage 3 AKIs are more likely to lead to a diagnostic code compared with stage 1 AKI).

**Table 4 | Advantages and disadvantages of different definitions of AKI used in previous studies based on routine care data**

| Definition | Advantages | Disadvantages |
|---|---|---|
| Diagnosis codes | - Available in settings without laboratory registries<br>- High specificity for severe AKI and AKI requiring dialysis | - Low sensitivity for AKI, especially for less severe stages<br>- Quality of coding relies on the specific health care setting, changes in diagnostic criteria, and coding practices over time<br>- AKI during elective admissions is less likely to be captured compared with admissions where AKI was the reason for hospitalization[78]<br>- Misclassification influenced by coding practices and purpose (e.g., reimbursement, pay for performance, and documentation in routine practice) |
| KDIGO serum creatinine criteria | - Possible to separate AKI from prevalent CKD when a valid baseline serum creatinine is available<br>- When definitions are harmonized, comparable standardized incidence rates of AKI across populations, allowing for direct comparison between studies | - Inpatient tests cannot distinguish AKI from preexisting CKD<br>- Outpatient tests may be missing<br>- Choice of numerous baseline serum creatinine definitions<br>- Sensitive to changes in testing practices |
| KDIGO urine output criteria | - Research indicates that short- and long-term risk of death or KRT is greatest when patients meet both serum creatinine and urine output criteria for AKI | - Seldom captured in administrative data, and rarely available outside the ICU |

AKI, acute kidney injury; CKD, chronic kidney disease; ICU, intensive care unit; KDIGO, Kidney Disease: Improving Global Outcomes; KRT, kidney replacement therapy.

When AKI is the outcome, using AKI diagnoses may impact the validity of findings. Whether bias occurs depends on the specificity and sensitivity of the outcome definition, as well as whether the interest lies in relative or absolute risks. When the specificity is high (i.e., the probability of not having AKI among those who truly do not have AKI equals 1 for both the exposed and unexposed), as is the case when using AKI diagnoses, relative risk estimates will not be biased, even if the sensitivity is low (i.e., probability of recorded AKI among those who truly have AKI).[81] However, absolute risk estimates will be biased, leading to an underestimation of the absolute risk difference of all AKI cases. As discussed earlier in the section on CKD progression, bias will also occur if the measurement error in the outcome is differential with respect to the exposure (e.g., in question 3, this will occur if physicians suspect that RASi use causes recurrent AKI and therefore monitor patients who continue RASi more closely than patients who stop RASi).[82] Such differential measurement error across exposure groups may be less likely for severe AKI (as ascertained by diagnosis codes), because these will be recorded regardless of exposure status.

When history of AKI is a confounder, as in the example question 3, using diagnosis codes may lead to residual confounding.[21] Whenever a patient had an AKI that was not severe enough to be coded, this measurement error leads to residual confounding when the prescriber was aware of the history of AKI, and bases his/her prescribing decision (stopping vs. continuing RASi) on this.

***Defining AKI cases by urine output.*** The 2012 KDIGO classification of AKI is currently widely used for both clinical and research purposes.[2] Using these criteria fully (i.e., considering both changes in serum creatinine and urine output) is recommended in clinical practice, because short- and long-term risk of death and KRT is greatest when patients meet both criteria.[83] However, this level of detail (e.g., hourly urine output) is not easily accessible in many routine health care databases, limiting their use in epidemiologic studies.[84,85] In most electronic health care records, urine output and point-of-care creatinine measurements are added to the electronic health records as unstructured text, which will hamper accurate extraction, although this problem may be mitigated by using natural language processing to extract and classify this information from unstructured texts. The same considerations regarding generalizability and bias as discussed above for AKI diagnosis codes apply for urine output.

***AKI based on creatinine.*** Although creatinine measurements are preferred to diagnosis codes, certain challenges arise when using routinely collected data sources. The KDIGO criteria for diagnosis of AKI in clinical practice refer to a relative increase in serum or plasma creatinine of $\geq$1.5, known or presumed to have occurred within the prior 7 days, or an absolute increase of $\geq$0.3 mg/dl ($\geq$26.5 μmol/l) within 48 hours. To avoid the influence of acute illness, outpatient serum creatinine tests are preferred sources to establish the baseline creatinine. Ideally, serum creatinine would need to have been measured within 7 days before AKI onset for detection of AKI.[62] However, this is seldom the case, except for situations such as planned surgeries. Researchers are then left with the option of applying different windows to identify prior serum creatinine measurements to define as to who had AKI.[62,86–89] Nevertheless, for a proportion of patients, a creatinine test within the specified period will be lacking.

A recent scoping review confirmed a lack of consistency in how KDIGO definitions for AKI were used in epidemiologic studies; for instance, the window to ascertain the baseline creatinine ranged from 0 days to more than a year before the AKI. More concerning, however, was the absence of description of the process used in 33% of the identified studies.[85]

If >1 eligible creatinine test is available per patient, it is unclear whether the preferred approach would be to select the most recent serum creatinine,[86,90,91] the median,[92,93] or the mean,[94] of all eligible tests, or to model the slope of creatinine and select its intersection with the 7-day period of interest. In 1 study, the mean outpatient serum creatinine measured in the year before hospitalization most closely approximated nephrologist-adjudicated "baseline" serum creatinine values.[62]

Because of lack of testing, most studies in the literature (71%) opt to exclude individuals who lack a baseline creatinine test.[85] Strategies used in previous literature to estimate baseline serum creatinine, when not measured, include simple or multiple imputation, using the serum creatinine at admission,[95] assuming an eGFR of 75 ml/min per 1.73 $m^2$,[96] or using a post-AKI nadir value.[97] Studies comparing these approaches suggest that multiple imputation is superior to simple imputation or assuming an eGFR of 75 ml/min per 1.73 $m^2$.[98] Using a nadir serum creatinine during hospitalization as baseline may lead to incorrect detection of AKI, because serum creatinine in the inpatient setting is influenced by nonrenal factors, such as fluid accumulation and loss of muscle mass.[99,100] Using the first serum creatinine on admission could result in AKI episodes being missed if serum creatinine was already elevated on admission. However, time lag between a kidney insult (due to an acute illness) and serum creatinine elevation should be acknowledged: it may take up to 48 to 72 hours after the kidney insult happened for creatinine to increase.[101] Indeed, a US study showed that the first inpatient serum creatinine was not higher than the most recent outpatient serum creatinine in a large proportion of hospitalized patients with AKI.[62] However, this may vary by cause of hospitalization. For example, in health systems with rapid admission for an acute ST-elevation myocardial infarction within hours of onset of chest pain, serum creatinine elevation would be only visible after admission. However, for admissions with infections and other conditions that gradually develop over several days, serum creatinine may be already elevated on admission.

Box 1 summarizes recommendations for clearly reporting the time frame for eligible baseline creatinine values and the rationale for doing so, how missing baseline creatinine values are handled, and the method chosen to select the baseline eGFR when there are multiple eligible values within the defining window, with reference to a recent consensus by a Delphi panel composed of nephrologists and epidemiologists with experience in AKI research.[85] A recent study showed that harmonizing AKI definitions across 4 population-based databases produced comparable standardized incidence rates of AKI.[102]

### Moving forward: toward more robust estimations

The longitudinal analysis of routinely collected health care data relies on the assumption that the timing and frequency of the measurement of longitudinal outcomes should be independent of the value of the outcome itself.[103–105] Understanding the extent to which this assumption is violated is important; patients will visit the physician when they have been feeling ill and hence have worse biomarker values; patients with comorbidities are likely to have more health care visits than patients without comorbidities. It becomes apparent that observations and outcomes are dependent, and thus missing laboratory tests are not completely at random. This has been referred to as "informative presence," or alternatively, "informative visit process," "dynamic observation plans,"[106] or "outcome-dependent visits" and is an aspect often ignored in research practice and can be considered a form of information bias.[107,108]

Relatively simple analyses can be performed to assess the magnitude of effects owing to informative visits in the data set. *First*, when the data set contains information on whether a visit is scheduled or unscheduled, the longitudinal eGFR slope can be calculated separately for scheduled and unscheduled visits. A substantial difference between slopes is suggestive of an informative visit process.[105] *Second*, one can calculate the correlation between a subject's eGFR value at a certain time point and the time between this measurement and the next, for all measurements.[105] Alternatively, the number of visits can be compared between individuals with a high or low eGFR. *Third*, when comparing 2 different interventions (e.g., question 2, dipeptidyl peptidase 4 inhibitors vs. sodium-glucose cotransporter-2 inhibitors), differential outcome ascertainment may be assessed by comparing the proportion of individuals with at least 1 creatinine measurement, the rate of creatinine testing during periods of treatment,[109,110] or the average time gap between tests. *Finally*, recurrent events models (such as the Andersen-Gill model) could be used to quantify the association between study covariates and the rate of observation.[111] Overall, simply reporting the number of visits per patient, gaps between visits, and potential predictors of visit time can give the reader an indication of the extent of irregularity and its informativeness.[107]

Some strategies may serve to mitigate the bias introduced by outcome-dependent visits; applying an active comparator design might yield a reference group with similar observation and dropout patterns, as described elsewhere, provided that testing rates are similar.[112–114] Bias can be attenuated when a certain proportion of the sample contains noninformative, regularly planned visits.[77,104,105,115,116] In many cohorts, at least part of the visits will be regular. If information is available on whether visits are planned or unplanned, the analysis could be restricted to preplanned visits to yield a cohort of subjects where the information process is independent of disease severity. Another option is to restrict the analytical sample to a population with an indication for regular kidney function monitoring (e.g., patients with diabetes),[33] at the expense of the external validity or generalizability of the study findings. The large sample sizes of health

**Box 1 | Reporting recommendations when studying AKI/AKD**

1. Studies should describe the intended target population (all patients with AKI? only diagnosed/severe AKI?), and whether study results are generalizable to that target population.
2. Studies should clarify how populations with/without baseline creatinine results differ (sample size, characteristics, and setting of testing [i.e., outpatient vs. inpatient results]), and the timing of the baseline creatinine relative to the AKI precipitating event.
3. Studies should clearly report the AKI definition used (e.g., whether 0.3 mg/dl increase over 48 hours is included [required for full alignment with the KDIGO AKI definition], whether staging criteria for stages 1, 2, and 3 are used, and whether urine output criteria were included).
4. Studies should clarify the definition of a baseline creatinine if multiple baseline creatinine results were available (e.g., was the mean of measurements used or the latest measurement? were measurements <7 days before AKI discarded?).
5. Studies should clearly report what was done whenever baseline results were not available. If studies impute missing baseline creatinine tests, they should specify methods used and discuss the implications of this imputation on study findings.

AKD, acute kidney disease; AKI, acute kidney injury; KDIGO, Kidney Disease: Improving Global Outcomes.

care data sets often allow for this type of selection, while still retaining power. Comparing results across different health systems with different testing indications is also helpful.

Alternatively, various methods have been proposed to accommodate an informative observation process or dropout in the study, such as, for example, due to referral to specialist care not covered by the database studied[103,117]; herein, we will discuss briefly approaches based on inverse probability weighting and approaches that aim to fully model the different processes. Methods based on inverse probability weighting rely on the idea of weighting each observation by the inverse probability of each measurement to be recorded or for the inverse probability of nondropout from the study; consequently, this approach creates a pseudopopulation in which the observation or the dropout process is static (rather than dynamic); that is, the process is completely at random and can, therefore, be ignored.[106,118,119] This can be implemented in practice using standard, off-the-shelf statistical software, or using user-contributed packages, such as the IrregLong package in R.[120]

Furthermore, it is possible to fully model the observation and/or dropout process within the joint longitudinal-survival modeling framework.[77,121–123] Joint models consist of 2 "submodels": 1 to model the survival outcome (i.e., the observation/dropout process), and the other to model the longitudinal outcome (i.e., the longitudinal kidney function measurements). A survival model is used to model the survival outcome, and a linear mixed model is used for the longitudinal outcome. The submodels are generally linked via shared random effects and estimated jointly. Focusing on dropout, the joint modeling approach can accommodate informative truncation of longitudinal trajectories due to dropout (e.g., death). Similar to linear mixed models, the baseline value should be part of the outcome vector as it contributes to estimating the measurement error. Compared with the inverse probability weighting approach, the joint modeling approach has the advantage of explicitly modeling all the processes of interest, allowing joint inference on the different aspects of the problem under study. However, this approach has the disadvantage of being more computationally intensive, limiting its applicability (especially with large data sets) as well as needing to specify the shared random effects correctly; the joint modeling approach can be implemented using readily available statistical software in R and Stata (e.g., the merlin package[124]).

## Future directions

Observational research of kidney disease has made great strides with studies including larger populations, more sophisticated

**Box 2 | Key points a causal study should consider discussing when using routinely collected health care data to study populations with CKD or CKD progression as an outcome**

- Investigate and discuss to what extent study results are generalizable to the target population in the context of the definition used to identify populations with CKD (e.g., based on diagnosis codes, eGFR measurements, and UACR measurements).
- Investigate and discuss the potential for differential outcome ascertainment (e.g., check whether more kidney function measurements are performed in 1 exposure group).
- Investigate the impact of exposure misclassification in the context of the definition used (diagnosis codes and eGFR based on serum creatinine).
- Discuss the key potential confounders of the exposure-outcome relationship, and discuss potential residual confounding (e.g., owing to disease severity or misclassification). Investigate the presence of residual confounding through positive or negative control outcomes, and its impact with quantitative bias analysis.
- When using eGFR measurements to classify CKD, discuss how multiple measurements are handled (mean, median, and most recent).
- When using eGFR or UACR as adjustment variables for confounding, discuss how patients with missing data were handled (complete case analysis, multiple imputation, and weighting). Discuss the possibility for selection bias when a complete case analysis is performed, also in light of the pattern of missingness and the proportion of patients with missing data.
- What are the data sources from which kidney function information for individuals was obtained? Does the chosen database adequately capture all kidney function measurements? Discuss consequences of data fragmentation on study results, including loss to follow-up.
- The use of a diagram is recommended to illustrate key aspects of the study design(s), including study entry, exposure, confirmation of exposure, comparison groups, lag and observation periods, and covariate definitions as relevant.

CKD, chronic pkidney disease; eGFR, estimated glomerular filtration rate; UACR, urine albumin–to–creatinine ratio.

analytic methods, individual-level meta-analysis, and sensitivity analyses to help gauge the validity of the results. However, as shown in this review, the field will benefit from more transparent and structured reporting, and thoughtful acknowledgement and discussion of potential biases.

Given that the suitability of each data set will depend on the research question and local structural factors, it may not be possible to impose a single strict definition that suits all studies that use routine health care data. However, validation studies are helpful within specific health systems to investigate the local sensitivity and specificity of these definitions. We advocate for concerted efforts to encourage improved reporting practices for routinely collected data on kidney exposures and outcomes.

The REporting of studies Conducted using Observational Routinely collected health Data for PharmacoEpidemiology (RECORD-PE) guidelines were produced as part of an international collaboration to improve such practice, building on the existing Strengthening the Reporting of Observational Studies in Epidemiology and RECORD guidelines.[17] We encourage researchers and editors of scientific journals to use this template to guide reporting of definitions of kidney exposures and outcomes (Supplementary Table S2). Key information that an article should address is outlined in Box 2.

Several specific approaches that are referred to are worth highlighting:

- Study design diagrams may serve to effectively illustrate algorithms used to define exposure or outcome for kidney function (including any sensitivity analyses).[125]
- Sensitivity analyses can be used to examine the assumptions underlying the chosen analytical approaches.
- Triangulation approaches to address biases can enhance causal inference.[126] This necessitates analyses in a range of settings and the integration of results from several approaches, each prone to their own different and unrelated sources of potential bias, to qualitatively determine and explicitly articulate the strength of evidence; examples include cross-context comparisons, such as different study populations, which would be expected to introduce their own inherent biases.
- Use of directed acyclic graph is recommended to consider bias, such as selection bias, and confounding.[127]
- Open working methods mandate open sharing of all analysis codes to encourage a culture of external review, reuse, and collaboration using a given source of data.[128] Similar approaches could be used for other large routinely collected data in other settings to enhance transparency and replication of analyses to enhance trust in research findings.

In conclusion, the perfect definition of kidney exposures, covariates, or outcomes using routinely collected data depends on the research question and availability of data, but clearer and more transparent reporting of these decisions in observational research is necessary to move the field forward.

## DISCLOSURE

All the authors declared no competing interests.

## ACKNOWLEDGMENTS

JJC acknowledges support from the Swedish Research Council. ELF acknowledges support by a Rubicon Grant of the Netherlands Organization for Scientific Research (NWO). MMS is funded by the Jindal Research Chair for the Prevention of Kidney Disease. CFC acknowledges support from the Independent Research Fund Denmark. VM is funded by the National Institutes for Health and Care Research (NIHR). FC is funded by the Medical Research Council (MR/N013638/1). JC acknowledges support from the National Institutes of Health and US National Kidney Foundation.

## SUPPLEMENTARY MATERIAL

Supplementary File (PDF)

**Table S1.** Fragmentation of care in different health care settings of the world.

**Table S2.** Adapted REporting of studies Conducted using Observational Routinely collected health Data for PharmacoEpidemiology (RECORD-PE) checklist for electronic health record studies involving measurement and/or classification of kidney function for exposure or outcome.

**Figure S1.** Directed acyclic graphs showing selection bias due to conditioning on a collider for the different example questions.

**Supplementary References.**

## REFERENCES

1. Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney Int Suppl*. 2013;3:1–150.
2. Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group. KDIGO clinical practice guideline for acute kidney injury. *Kidney Int Suppl*. 2012;2:1–138.
3. Chawla LS, Bellomo R, Bihorac A, et al. Acute kidney disease and renal recovery: consensus report of the Acute Disease Quality Initiative (ADQI) 16 Workgroup. *Nat Rev Nephrol*. 2017;13:241–257.
4. Lameire NH, Levin A, Kellum JA, et al. Harmonizing acute and chronic kidney disease definition and classification: report of a Kidney Disease: Improving Global Outcomes (KDIGO) Consensus Conference. *Kidney Int*. 2021;100:516–526.
5. Stevens LA, Greene T, Levey AS. Surrogate end points for clinical trials of kidney disease progression. *Clin J Am Soc Nephrol*. 2006;1:874–884.
6. Perkovic V, Koitka-Weber A, Cooper ME, et al. Choice of endpoint in kidney outcome trials: considerations from the EMPA-REG OUTCOME® trial. *Nephrol Dial Transplant*. 2020;35:2103–2111.
7. Levin A, Agarwal R, Herrington WG, et al. International consensus definitions of clinical trial outcomes for kidney failure: 2020. *Kidney Int*. 2020;98:849–859.
8. Standardised Outcomes in Nephrology (SONG). The SONG initiative. Accessed November 10, 2021. https://songinitiative.org/
9. Shrier I, Stovitz SD, Wang C, Steele RJ. Beware of collider stratification bias when analyzing recurrent injuries. *Scand J Med Sci Sports*. 2022;32:270–272.
10. Logan R, Davey P, De Souza N, et al. Assessing the accuracy of *ICD-10* coding for measuring rates of and mortality from acute kidney injury and the impact of electronic alerts: an observational cohort study. *Clin Kidney J*. 2020;13:1083–1090.
11. Hwang YJ, Shariff SZ, Gandhi S, et al. Validity of the *International Classification of Diseases, Tenth Revision* code for acute kidney injury in elderly patients at presentation to the emergency department and at hospital admission. *BMJ Open*. 2012;2:e001821.
12. Waikar SS, Wald R, Chertow GM, et al. Validity of *International Classification of Diseases, Ninth Revision, Clinical Modification* codes for acute renal failure. *J Am Soc Nephrol*. 2006;17:1688–1694.
13. Grams ME, Waikar SS, MacMahon B, et al. Performance and limitations of administrative data in the identification of AKI. *Clin J Am Soc Nephrol*. 2014;9:682–689.
14. Last JM. *A Dictionary of Epidemiology*. 4th ed. Oxford University Press; 2000.
15. Heinze G, Wallisch C, Kainz A, et al. Chances and challenges of using routine data collections for renal health care research. *Nephrol Dial Transplant*. 2015;30(suppl 4):iv68–iv75.

16. Fu EL, Evans M, Carrero JJ, et al. Timing of dialysis initiation to reduce mortality and cardiovascular events in advanced chronic kidney disease: nationwide cohort study. *BMJ*. 2021;375:e066306.
17. Langan SM, Schmidt SA, Wing K, et al. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). *BMJ*. 2018;363:k3532.
18. Toh ST, Andrade SE, Raebel MA, et al. Examples of existing automated databases. In: Strom BL, Kimmel SE, Hennessy S, eds. *Textbook of Pharmacoepidemiology*. 2nd ed. Wiley Blackwell; 2013:123–177.
19. Gerhard T, Moride Y, Pottegård A, Pratt N. Encounter databases. In: Strom BL, Kimmel SE, Hennessy S, eds. *Pharmacoepidemiology*. 6th ed. Wiley Blackwell; 2019:211–240.
20. Wang SV, Pinheiro S, Hua W, et al. STaRT-RWE: structured template for planning and reporting on the implementation of real world evidence studies. *BMJ*. 2021;372:m4856.
21. Funk MJ, Landi SN. Misclassification in administrative claims data: quantifying the impact on treatment effect estimates. *Curr Epidemiol Rep*. 2014;1:175–185.
22. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15:615–625.
23. Runesson B, Gasparini A, Qureshi AR, et al. The Stockholm CREAtinine Measurements (SCREAM) project: protocol overview and regional representativeness. *Clin Kidney J*. 2016;9:119–127.
24. Chesnaye NC, Stel VS, Tripepi G, et al. An introduction to inverse probability of treatment weighting in observational research. *Clin Kidney J*. 2022;15:14–20.
25. Shin JI, Chang AR, Grams ME, et al. Albuminuria testing in hypertension and diabetes: an individual-participant data meta-analysis in a global consortium. *Hypertension*. 2021;78:1042–1052.
26. Bello AK, Ronksley PE, Tangri N, et al. Quality of chronic kidney disease management in Canadian primary care. *JAMA Netw Open*. 2019;2:e1910704.
27. Nitsch D, Caplin B, Hull S, Wheeler DC. National chronic kidney disease audit national report. Accessed November 10, 2021. https://www.lshtm.ac.uk/files/ckd_audit_report.pdf
28. Fu EL, Clase CM, Evans M, et al. Comparative effectiveness of renin-angiotensin system inhibitors and calcium channel blockers in individuals with advanced CKD: a nationwide observational cohort study. *Am J Kidney Dis*. 2021;77:719–729.e1.
29. Hughes RA, Heron J, Sterne JAC, Tilling K. Accounting for missing data in statistical analyses: multiple imputation is not always the answer. *Int J Epidemiol*. 2019;48:1294–1304.
30. Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009;338:b2393.
31. Madley-Dowd P, Hughes R, Tilling K, Heron J. The proportion of missing data should not be used to guide decisions on multiple imputation. *J Clin Epidemiol*. 2019;110:63–73.
32. Blazek K, van Zwieten A, Saglimbene V, Teixeira-Pinto A. A practical guide to multiple imputation of missing data in nephrology. *Kidney Int*. 2021;99:68–74.
33. McDonald HI, Shaw C, Thomas SL, et al. Methodological challenges when carrying out research on CKD and AKI using routine electronic health records. *Kidney Int*. 2016;90:943–949.
34. Anderson J, Glynn LG. Definition of chronic kidney disease and measurement of kidney function in original research papers: a review of the literature. *Nephrol Dial Transplant*. 2011;26:2793–2798.
35. Paik JM, Patorno E, Zhuo M, et al. Accuracy of identifying diagnosis of moderate to severe chronic kidney disease in administrative claims data. *Pharmacoepidemiol Drug Saf*. 2022;31:467–475.
36. Gasparini A, Evans M, Coresh J, et al. Prevalence and recognition of chronic kidney disease in Stockholm healthcare. *Nephrol Dial Transplant*. 2016;31:2086–2094.
37. Vestergaard SV, Christiansen CF, Thomsen RW, et al. Identification of patients with CKD in medical databases: a comparison of different algorithms. *Clin J Am Soc Nephrol*. 2021;16:543–551.
38. Tuttle KR, Alicic RZ, Duru OK, et al. Clinical characteristics of and risk factors for chronic kidney disease among adults and children: an analysis of the CURE-CKD Registry. *JAMA Netw Open*. 2019;2:e1918169.
39. van Oosten MJM, Brohet RM, Logtenberg SJJ, et al. The validity of Dutch health claims data for identifying patients with chronic kidney disease: a hospital-based study in the Netherlands. *Clin Kidney J*. 2020;14:1586–1593.
40. van Oosten MJM, SJJ Logtenberg, Edens MA, et al. Health claims databases used for kidney research around the world. *Clin Kidney J*. 2021;14:84–97.
41. Roy L, Zappitelli M, White-Guay B, et al. Agreement between administrative database and medical chart review for the prediction of chronic kidney disease G category. *Can J Kidney Health Dis*. 2020;7:2054358120959908.
42. Castro AF, Coresh J. CKD surveillance using laboratory data from the population-based National Health and Nutrition Examination Survey (NHANES). *Am J Kidney Dis*. 2009;53(suppl 3):S46–S55.
43. Brück K, Jager KJ, Dounousi E, et al. Methodology used in studies reporting chronic kidney disease prevalence: a systematic literature review. *Nephrol Dial Transplant*. 2015;30(suppl 4):iv6–iv16.
44. Tuot DS, McCulloch CE, Velasquez A, et al. Impact of a primary care CKD registry in a US public safety-net health care delivery system: a pragmatic randomized trial. *Am J Kidney Dis*. 2018;72:168–177.
45. de Boer IH, Kovesdy CP, Navaneethan SD, et al. Pragmatic clinical trials in CKD: opportunities and challenges. *J Am Soc Nephrol*. 2016;27:2948–2954.
46. Navaneethan SD, Jolly SE, Sharp J, et al. Electronic health records: a new tool to combat chronic kidney disease? *Clin Nephrol*. 2013;79:175–183.
47. Navaneethan SD, Jolly SE, Schold JD, et al. Development and validation of an electronic health record-based chronic kidney disease registry. *Clin J Am Soc Nephrol*. 2011;6:40–49.
48. Hernan MA, Sauer BC, Hernandez-Diaz S, et al. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol*. 2016;79:70–75.
49. Shariff SZ, Cuerden MS, Jain AK, Garg AX. The secret of immortal time bias in epidemiologic studies. *J Am Soc Nephrol*. 2008;19:841–843.
50. Tam-Tham H, Quinn RR, Weaver RG, et al. Survival among older adults with kidney failure is better in the first three years with chronic dialysis treatment than not. *Kidney Int*. 2018;94:582–588.
51. Fu EL, van Diepen M. Comment on Kwon et al: the long-term effects of metformin on patients with type 2 diabetic kidney disease: *Diabetes Care* 2020;43:948–955. *Diabetes Care*. 2020;43:e190.
52. Stovitz SD, Banack HR, Kaufman JS. "Depletion of the susceptibles" taught through a story, a table and basic arithmetic. *BMJ Evid Based Med*. 2018;23:199.
53. Fu EL, Evans M, Clase CM, et al. Stopping renin-angiotensin system inhibitors in patients with advanced CKD and risk of adverse outcomes: a nationwide study. *J Am Soc Nephrol*. 2021;32:424–435.
54. Hernan MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol*. 2016;183:758–764.
55. Delanaye P, Glassock RJ, De Broe ME. Epidemiology of chronic kidney disease: think (at least) twice! *Clin Kidney J*. 2017;10:370–374.
56. Ocak G, Khairoun M, Khairoun O, et al. Chronic kidney disease and atrial fibrillation: a dangerous combination. *PLoS One*. 2022;17:e0266046.
57. Hill NR, Fatoba ST, Oke JL, et al. Global prevalence of chronic kidney disease - a systematic review and meta-analysis. *PLoS One*. 2016;11:e0158765.
58. Sumida K, Nadkarni GN, Grams ME, et al. Conversion of urine protein-creatinine ratio or urine dipstick protein to urine albumin-creatinine ratio for use in chronic kidney disease screening and prognosis: an individual participant-based meta-analysis. *Ann Intern Med*. 2020;173:426–435.
59. Weaver RG, James MT, Ravani P, et al. Estimating urine albumin-to-creatinine ratio from protein-to-creatinine ratio: development of equations using same-day measurements. *J Am Soc Nephrol*. 2020;31:591–601.
60. White SL, Yu R, Craig JC, et al. Diagnostic accuracy of urine dipsticks for detection of albuminuria in the general community. *Am J Kidney Dis*. 2011;58:19–28.
61. Mendu ML, Ahmed S, Maron JK, et al. Development of an electronic health record-based chronic kidney disease registry to promote population health management. *BMC Nephrol*. 2019;20:72.
62. Siew ED, Ikizler TA, Matheny ME, et al. Estimating baseline kidney function in hospitalized patients with impaired kidney function. *Clin J Am Soc Nephrol*. 2012;7:712–729.
63. Pasternak B, Wintzell V, Melbye M, et al. Use of sodium-glucose co-transporter 2 inhibitors and risk of serious renal events: Scandinavian cohort study. *BMJ*. 2020;369:m1186.
64. Hundemer GL, Knoll GA, Petrcich W, et al. Kidney, cardiac, and safety outcomes associated with alpha-blockers in patients with CKD: a population-based cohort study. *Am J Kidney Dis*. 2021;77:178–189.e1.

65. Fu EL, Franko MA, Obergfell A, et al. High-sensitivity C-reactive protein and the risk of chronic kidney disease progression or acute kidney injury in post-myocardial infarction patients. *Am Heart J*. 2019;216:20–29.
66. Novak M, Mucsi I, Rhee CM, et al. Increased risk of incident chronic kidney disease, cardiovascular disease, and mortality in patients with diabetes with comorbid depression. *Diabetes Care*. 2016;39:1940–1947.
67. Wetmore JB, Yan H, Herzog CA, et al. CKD progression in Medicare beneficiaries with nonvalvular atrial fibrillation treated with apixaban versus warfarin. *Am J Kidney Dis*. 2021;78:180–189.
68. Hodlmoser S, Carrero JJ, Kurnikowski A, et al. Kidney function, kidney replacement therapy, and mortality in men and women. *Kidney Int Rep*. 2022;7:444–454.
69. Levey AS, Inker LA, Matsushita K, et al. GFR decline as an end point for clinical trials in CKD: a scientific workshop sponsored by the National Kidney Foundation and the US Food and Drug Administration. *Am J Kidney Dis*. 2014;64:821–835.
70. Levey AS, Gansevoort RT, Coresh J, et al. Change in albuminuria and GFR as end points for clinical trials in early stages of CKD: a scientific workshop sponsored by the National Kidney Foundation in collaboration with the US Food and Drug Administration and European Medicines Agency. *Am J Kidney Dis*. 2020;75:84–104.
71. Ragland DR. Dichotomizing continuous outcome variables: dependence of the magnitude of association and statistical power on the cutpoint. *Epidemiology*. 1992;3:434–440.
72. Neuen BL, Weldegiorgis M, Herrington WG, et al. Changes in GFR and albuminuria in routine clinical practice and the risk of kidney disease progression. *Am J Kidney Dis*. 2021;78:350–360.e1.
73. Zee J, Mansfield S, Mariani LH, Gillespie BW. Using all longitudinal data to define time to specified percentages of estimated GFR decline: a simulation study. *Am J Kidney Dis*. 2019;73:82–89.
74. Leffondre K, Boucquemont J, Tripepi G, et al. Analysis of risk factors associated with renal function trajectory over time: a comparison of different statistical approaches. *Nephrol Dial Transplant*. 2015;30:1237–1243.
75. Heinze G, Christensen J, Haller MC. Modeling pulse wave velocity trajectories-challenges, opportunities, and pitfalls. *Kidney Int*. 2022;101:459–462.
76. Janmaat CJ, van Diepen M, Tsonaka R, et al. Pitfalls of linear regression for estimating slopes over time and how to avoid them by using linear mixed-effects models. *Nephrol Dial Transplant*. 2019;34:561–566.
77. Gasparini A, Abrams KR, Barrett JK, et al. Mixed-effects models for health care longitudinal data with an informative visiting process: a Monte Carlo simulation study. *Stat Neerl*. 2020;74:5–23.
78. Savino M, Plumb L, Casula A, et al. Acute kidney injury identification for pharmacoepidemiologic studies: use of laboratory electronic acute kidney injury alerts versus electronic health records in hospital episode statistics. *Pharmacoepidemiol Drug Saf*. 2021;30:1687–1695.
79. Janse RJ, Fu EL, Clase CM, et al. Stopping versus continuing renin–angiotensin–system inhibitors after acute kidney injury and adverse clinical outcomes: an observational study from routine care data. *Clin Kidney J*. 2022;15:1109–1119.
80. James MT, Levey AS, Tonelli M, et al. Incidence and prognosis of acute kidney diseases and disorders using an integrated approach to laboratory measurements in a universal health care system. *JAMA Netw Open*. 2019;2:e191795.
81. Lash TL, Fox MP, Fink AK. *Applying Quantitative Bias Analysis to Epidemiologic Data*. Springer; 2009.
82. Bidulka P, Fu EL, Leyrat C, et al. Stopping renin-angiotensin system blockers after acute kidney injury and risk of adverse outcomes: parallel population-based cohort studies in English and Swedish routine care. *BMC Med*. 2020;18:195.
83. Kellum JA, Sileanu FE, Murugan R, et al. Classifying AKI by urine output versus serum creatinine level. *J Am Soc Nephrol*. 2015;26:2231–2238.
84. Levey AS, Eckardt KU, Dorman NM, et al. Nomenclature for kidney function and disease: report of a Kidney Disease: Improving Global Outcomes (KDIGO) Consensus Conference. *Kidney Int*. 2020;97:1117–1129.
85. Guthrie G, Guthrie B, Walker H, et al. Developing an AKI consensus definition for database research: findings from a scoping review and expert opinion using a Delphi process. *Am J Kidney Dis*. 2022;79:488–496.e1.
86. Liu KD, Hsu CY, Yang J, et al. Acute kidney injury ascertainment is affected by the use of first inpatient versus outpatient baseline serum creatinine. *Kidney Int Rep*. 2018;3:211–215.
87. Cooper DJ, Plewes K, Grigg MJ, et al. An evaluation of commonly used surrogate baseline creatinine values to classify AKI during acute infection. *Kidney Int Rep*. 2021;6:645–656.
88. Bouchard J. Estimating baseline serum creatinine for assessing acute kidney injury: not a one size fits all approach. *Kidney Int Rep*. 2021;6:562–564.
89. Flynn N. Estimating baseline kidney function in hospitalized patients with impaired kidney function. *Ann Clin Biochem*. 2013;50:89.
90. Hsu CY, Chinchilli VM, Coca S, et al. Post-acute kidney injury proteinuria and subsequent kidney disease progression: the Assessment, Serial Evaluation, and Subsequent Sequelae in Acute Kidney Injury (ASSESS-AKI) Study. *JAMA Intern Med*. 2020;180:402–410.
91. Siew ED, Peterson JF, Eden SK, et al. Outpatient nephrology referral rates after acute kidney injury. *J Am Soc Nephrol*. 2012;23:305–312.
92. Sawhney S, Marks A, Fluck N, et al. Intermediate and long-term outcomes of survivors of acute kidney injury episodes: a large population-based cohort study. *Am J Kidney Dis*. 2017;69:18–28.
93. Kashani K, Shao M, Li G, et al. No increase in the incidence of acute kidney injury in a population-based annual temporal trends epidemiology study. *Kidney Int*. 2017;92:721–728.
94. Siew ED, Parr SK, Abdel-Kader K, et al. Predictors of recurrent AKI. *J Am Soc Nephrol*. 2016;27:1190–1200.
95. Liaño F, Pascual J, Madrid Acute Renal Failure Study Group. Epidemiology of acute renal failure: a prospective, multicenter, community-based study. *Kidney Int*. 1996;50:811–818.
96. Bagshaw SM, Uchino S, Cruz D, et al. A comparison of observed versus estimated baseline creatinine for determination of RIFLE class in patients with acute kidney injury. *Nephrol Dial Transplant*. 2009;24:2739–2744.
97. Hsu RK, Hsu CY, McCulloch CE, et al. Research-based versus clinical serum creatinine measurements and the association of acute kidney injury with subsequent kidney function: findings from the Chronic Renal Insufficiency Cohort study. *Clin Kidney J*. 2020;13:55–62.
98. Siew ED, Peterson JF, Eden SK, et al. Use of multiple imputation method to improve estimation of missing baseline serum creatinine in acute kidney injury research. *Clin J Am Soc Nephrol*. 2013;8:10–18.
99. Prowle JR, Kolic I, Purdell-Lewis J, et al. Serum creatinine changes associated with critical illness and detection of persistent renal dysfunction after AKI. *Clin J Am Soc Nephrol*. 2014;9:1015–1023.
100. Liu KD, Thompson BT, Ancukiewicz M, et al. Acute kidney injury in patients with acute lung injury: impact of fluid accumulation on classification of acute kidney injury and associated outcomes. *Crit Care Med*. 2011;39:2665–2671.
101. Koyner JL. Assessment and diagnosis of renal dysfunction in the ICU. *Chest*. 2012;141:1584–1594.
102. Sawhney S, Bell S, Black C, et al. Harmonization of epidemiology of acute kidney injury and acute kidney disease produces comparable findings across four geographic populations. *Kidney Int*. 2022;101:1271–1281.
103. Pullenayegum EM, Lim LS. Longitudinal data subject to irregular observation: a review of methods with a focus on visit processes, assumptions, and study design. *Stat Methods Med Res*. 2016;25:2992–3014.
104. Neuhaus JM, McCulloch CE, Boylan RD. Analysis of longitudinal data from outcome-dependent visit processes: failure of proposed methods in realistic settings and potential improvements. *Stat Med*. 2018;37:4457–4471.
105. McCulloch CE, Neuhaus JM, Olin RL. Biased and unbiased estimation in longitudinal studies with informative visit processes. *Biometrics*. 2016;72:1315–1324.
106. Hernan MA, McAdams M, McGrath N, et al. Observation plans in longitudinal studies with time-varying treatments. *Stat Methods Med Res*. 2009;18:27–52.
107. Farzanfar D, Abumuamar A, Kim J, et al. Longitudinal studies that use data collected as part of usual care risk reporting biased results: a systematic review. *BMC Med Res Methodol*. 2017;17:133.
108. McGee G, Haneuse S, Coull BA, et al. On the nature of informative presence bias in analyses of electronic health records. *Epidemiology*. 2022;33:105–113.
109. Iskander C, Cherney DZ, Clemens KK, et al. Use of sodium-glucose cotransporter-2 inhibitors and risk of acute kidney injury in older adults

with diabetes: a population-based cohort study. *CMAJ*. 2020;192:E351–E360.
110. Xu Y, Fu EL, Clase CM, et al. GLP-1 receptor agonist versus DPP-4 inhibitor and kidney and cardiovascular outcomes in clinical practice in type-2 diabetes. *Kidney Int*. 2022;101:360–368.
111. Amorim LD, Cai J. Modelling recurrent events: a tutorial for analysis in epidemiology. *Int J Epidemiol*. 2015;44:324–333.
112. Rassen JA, Murk W, Schneeweiss S. Real-world evidence of bariatric surgery and cardiovascular benefits using electronic health records data: a lesson in bias. *Diabetes Obes Metab*. 2021;23:1453–1462.
113. Fu EL, van Diepen M, Xu Y, et al. Pharmacoepidemiology for nephrologists (part 2): potential biases and how to overcome them. *Clin Kidney J*. 2021;14:1317–1326.
114. Yoshida K, Solomon DH, Kim SC. Active-comparator design and new-user design in observational studies. *Nat Rev Rheumatol*. 2015;11:437–441.
115. Geskus RB. Which individuals make dropout informative? *Stat Methods Med Res*. 2014;23:91–106.
116. Goldstein BA, Phelan M, Pagidipati NJ, Peskoe SB. How and when informative visit processes can bias inference when using electronic health records data for clinical research. *J Am Med Inform Assoc*. 2019;26:1609–1617.
117. Sisk R, Lin L, Sperrin M, et al. Informative presence and observation in routine health data: a review of methodology for clinical risk prediction. *J Am Med Inform Assoc*. 2021;28:155–166.
118. Kurland BF, Heagerty PJ. Directly parameterized regression conditioning on being alive: analysis of longitudinal data truncated by deaths. *Biostatistics*. 2005;6:241–258.
119. Buzkova P, Brown ER, John-Stewart GC. Longitudinal data analysis for generalized linear models under participant-driven informative follow-up: an application in maternal health epidemiology. *Am J Epidemiol*. 2010;171:189–197.
120. Pullenayegum E. IrregLong: analysis of longitudinal data with irregular observation times: R package version 0.3.3. Accessed November 10, 2021. https://CRAN.R-project.org/package=IrregLong
121. Lawrence Gould A, Boye ME, Crowther MJ, et al. Joint modeling of survival and longitudinal non-survival data: current methods and issues: report of the DIA Bayesian joint modeling working group. *Stat Med*. 2015;34:2181–2195.
122. Harhay MO, Gasparini A, Walkey AJ, et al. Assessing the course of organ dysfunction using joint longitudinal and time-to-event modeling in the Vasopressin and Septic Shock Trial. *Crit Care Explor*. 2020;2:e0104.
123. Chesnaye NC, Tripepi G, Dekker FW, et al. An introduction to joint models-applications in nephrology. *Clin Kidney J*. 2020;13:143–149.
124. Crowther MJ. merlin—A unified modeling framework for data analysis and methods development in Stata. *Stata J*. 2020;20:763–784.
125. Schneeweiss S, Rassen JA, Brown JS, et al. Graphical depiction of longitudinal study designs in health care databases. *Ann Intern Med*. 2019;170:398–406.
126. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. *Int J Epidemiol*. 2016;45:1866–1886.
127. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10:37–48.
128. Williamson EJ, Walker AJ, Bhaskaran K, et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature*. 2020;584:430–436.