Methodological challenges when carrying out research on CKD and AKI using routine electronic health records

OPEN

Helen I. McDonald[1,3], Catriona Shaw[2,3], Sara L. Thomas[1], Kathryn E. Mansfield[1], Laurie A. Tomlinson[1] and Dorothea Nitsch[1]

[1]Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK; and [2]Renal Unit, King's College Hospital, London, UK

Research regarding chronic kidney disease (CKD) and acute kidney injury (AKI) using routinely collected data presents particular challenges. The availability, consistency, and quality of renal data in electronic health records has changed over time with developments in policy, practice incentives, clinical knowledge, and associated guideline changes. Epidemiologic research may be affected by patchy data resulting in an unrepresentative sample, selection bias, misclassification, and confounding by factors associated with testing for and recognition of reduced kidney function. We systematically explore the issues that may arise in study design and interpretation when using routine data sources for CKD and AKI research. First, we discuss how access to health care and management of patients with CKD may have an impact on defining the target population for epidemiologic study. We then consider how testing and recognition of CKD and AKI may lead to biases and how to potentially mitigate against these. Illustrative examples from our own research within the UK are used to clarify key points. Any research using routine renal data has to consider the local clinical context to achieve meaningful interpretation of the study findings.

Kidney International (2016) ■, ■–■; http://dx.doi.org/10.1016/j.kint.2016.04.010

KEYWORDS: electronic health records; epidemiological methods; renal disease

Correspondence: *Dorothea Nitsch, London School of Hygiene and Tropical Medicine, Faculty of Epidemiology and Population Health, Keppel Street, London, England WC1E 7HT, UK. E-mail: Dorothea.Nitsch@lshtm.ac.uk*

[3]These authors contributed equally to this work.

Received 5 February 2016; revised 7 April 2016; accepted 12 April 2016

Acute kidney injury (AKI) and chronic kidney disease (CKD) each have a substantial global health care burden.[1] The elderly with multiple morbidities are at most risk, and the burden of both diseases is set to rise in line with the demographic transition to older populations.[2] There is very limited understanding on how best to manage patients with multiple health conditions.[3] Historically, epidemiological research on this group was challenging. Large cohorts with many years of follow-up and high rates of retention were needed, resulting in logistical complexities and substantial financial costs. This has changed with the availability of computerized health care records. Data from electronic health records (EHRs) have facilitated large epidemiological studies to answer important questions that would otherwise not have been possible.[4]

Specific characteristics of kidney disease can present challenges when using routine data for research. Kidney disease is often asymptomatic, and diagnosis relies on blood and urine tests. There are wide disparities in the availability of records of renal function in EHRs.[5] The patchy nature of renal data means that epidemiologic research may be affected by biases. There is increasing use of EHR data in research and as part of performance management in health care. Recently, a reporting guideline for observational studies of routinely collected health care data was published, but no specific guidance exists how to apply this guideline in the context of renal research.[6]

The aim of this mini review is to systematically explore the issues that may arise in study design and interpretation of kidney disease in routine data. We focus on primary care data, as most patients with CKD are seen in the community setting. We will not discuss ethical issues. An excellent in-depth discussion of AKI research using secondary care data has been provided by others.[7]

In general, when reviewing EHR studies, it is useful to consider what the perfect study to address the question would be. Who should the perfect study focus on, and which data items would be needed to control for confounding? Contrasting this perfect scenario with the reality of the databases then clarifies the main limitations and potential biases.

Who is captured in the database?

How patients access health care impacts on who is recorded within routine data. For renal patients in particular, this may

reflect the cause and severity of their renal disease. For example, patients with early CKD are more likely to be identified and monitored in the primary care setting if they have known risk factors such as diabetes mellitus, while those receiving dialysis and under specialist care may only be recorded in secondary care datasets. Some health systems incentivize routine health checks (e.g., in occupational settings), which may include kidney function markers. If appropriate ethics approvals are in place, patient identifiers can be used to collect further information from other routine health data, such as hospitalization or dialysis registry data, to investigate long-term outcomes according to baseline function (examples given in Table 1).

In some settings, primary care practitioners are the gatekeepers for access to specialized care. Therefore, in theory, these data should be complete for every patient even without linkage. However, this assumes good information flow between primary and secondary care, and accurate recording of secondary/tertiary care episodes. Whether this is indeed the case is often unknown: for example, there are situations where patients may be able to access specialist care directly without first seeing a primary care practitioner (e.g., patients on renal replacement therapy). In health insurance–financed settings, the most complete data on individual patient care may be found in medical claims data held by the insurance company. However, laboratory results will often not be available within these data.

Table 1 | UK example: Potential sources of anonymized information about kidney disease

Primary care computerized health records

Databases are provided from specific primary care software providers; contain information on clinical diagnoses, prescriptions, medical procedures, and laboratory tests; and are traditionally coded with the Read clinical coding system and include feedback from secondary care. Examples include the Clinical Practice Research Datalink (CPRD) and The Health Improvement Network (THIN).

Hospital records

Hospital Episode Statistics (HES) provide information on diagnoses (coded with the *International Classification of Diseases, Tenth Revision* [*ICD-10*]) and medical procedures related to all National Health Service–funded hospital admissions in England. These data are entered by coders who look at notes and discharge letters. They do not contain information on laboratory results or inpatient prescriptions. Similar data to HES are collected in Wales, Northern Ireland, and Scotland. HES data can be linked to primary care records to provide more complete patient information.

Laboratory records

It is possible to extract data directly from laboratories in the UK, as is being done for acute kidney injury detection by the National Think Kidney's program. These data lack detailed patient information such as underlying diagnoses or prescriptions.

Pharmacy dispensing records

Such data may provide valuable information on whether a patient collected a prescribed drug from the pharmacy. However, there is little information on comorbidity and no information on laboratory test results. Data are usually linked to data containing information on renal patients.

Disease registries and audit

As part of National Audit used for monitoring the quality of care and commissioning, a range of disease registries have been set up to capture key features of routine clinical care for specific disease entities. This includes the UK Renal Registry, the National CKD Audit, and other registries that may collect some renal data (e.g., Diabetes Registry).

Initiation of renal replacement therapy is often used as a proxy for kidney failure (estimated glomerular filtration rate [eGFR] <15 ml/min/1.73 m^2) or end-stage renal disease. However, it is important to understand the characteristics of people who do not have access to care or are not recorded in selected EHRs. For example, in some health-care settings, patients have to pay for renal replacement therapy, for example, in South Africa $<50\%$ of patients with end-stage renal disease were accepted onto dialysis because of factors related to poverty.[8,9] Following the example of the US registry, many registries define dialysis continuing beyond 3 months as "chronic dialysis." The UK renal registry initially only collected data on patients who started and stayed on dialysis for at least 3 months; consequently, patients with end-stage renal disease who initiated dialysis but died before 3 months elapsed were missed. Patients may not want to start renal replacement therapy and instead undergo conservative management, but most renal registries do not collect this information. In Canada, among those aged 85 years or older, the proportion of untreated kidney failure was estimated to be up to 10-fold higher than those treated.[10]

In order to avoid bias, only patients expected to be tested as part of routine care for baseline kidney function and who would have outcomes of interest recorded if these happened should form part of the target population for study. Having identified this target population of interest, the next step is to check what information can be readily extracted and whether the research question of interest can be addressed.

Who had renal function tests and why?

Testing is often triggered by acute illness and therefore will not reflect "baseline" renal function. This differs from large epidemiological studies that include patients who volunteer to participate and typically are not acutely unwell. However, particularly in the older population in whom most routine tests are done, low eGFR measurements are usually not transient.[11]

Testing may also be influenced by the clinical context, for example, guidance recommends targeted testing for CKD in those at most risk (people with diabetes, hypertension, and cardiovascular disease).[12] In the UK, in response to guideline changes (Figure 1), there was an increase in recording of serum creatinine over time in a cohort of people with diabetes in primary care (Figure 2). Hence, the characteristics of patients who had renal function tested in the late 1990s may be very different from patients routinely tested from 2007/2008 onward.

For CKD—should codes or original laboratory data be used?

There is a distinction to be made between having recorded creatinine test results and a coded diagnosis of CKD. A coded diagnosis relies on the general practitioner's knowledge about CKD as well as incentives to formally code the diagnosis. There is wide variation in coding, with on average only 50%

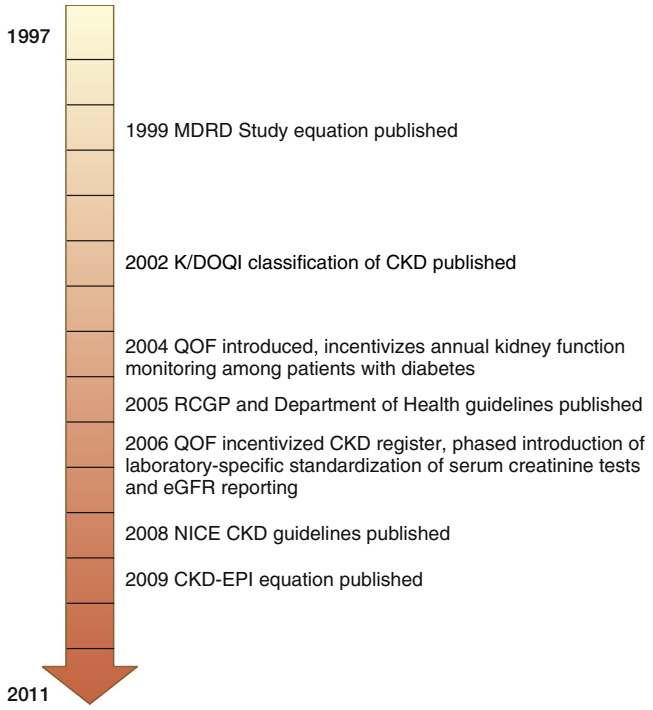

Figure 1 | UK example: timeline of changes to the identification of CKD in UK primary care during 1997–2011.[30] CKD-EPI, Chronic Kidney Disease Epidemiology Collaboration; K/DOQI, Kidney Disease Outcomes Quality Initiative; MDRD, Modification of Diet in Renal Disease; NICE, National Institute for Health and Care Excellence; QOF, Quality and Outcomes Framework; RCGP, Royal College of General Practitioners.

of patients meeting biochemical criteria for CKD being coded as such.[13] Defining patient groups using CKD codes alone results in the potential to miss a proportion of relevant patients to a variable extent over time.[14–17] Therefore, most large-scale epidemiologic research in EHRs relies on serum creatinine results.

Techniques for laboratory measurement of serum creatinine have changed over time. In the UK, laboratory variation in measurement of serum creatinine was addressed with the phased introduction of laboratory-specific standardization from 2006. Over time, laboratories have started to calibrate creatinine assays to a reference assay using isotope-dilution mass spectrometry, which needs to be considered in analyses. Therefore, eGFR calculated from serum creatinine levels in EHRs may need adjustment for changes in creatinine calibration depending on the time period covered in a study.[18] If reported eGFR results are used, care needs to be taken on which estimation formula (Modification of Diet in Renal Disease [MDRD], Chronic Kidney Disease Epidemiology Collaboration [CKD-EPI]) was used for reporting.

Proteinuric kidney disease is generally not well recorded, and there is substantial misclassification because positive results are more likely to be encoded than negative results.[5,15,16] For example, proteinuria levels detected on urinary test strips may be recorded as free text in the patient's medical notes but not formally coded in the EHRs, and therefore the information is harder to obtain. In routine care, more sensitive and better quantified tests such as the urinary albumin–creatinine ratio are usually limited to targeted populations known to be at high risk such as those with diabetes.[15] Previous studies using routinely collected data have used single time point measures of urinary protein or albuminuria, often incorporating different methods of quantification.[4,19] Most studies using outpatient laboratory data will not be able to quantify the precise timing of when the sample was taken (e.g., morning urine versus random urine).

For AKI— should codes or original laboratory data be used?

A universal definition and staging system for AKI was first introduced in 2004.[20] Its aim was to present AKI as a spectrum, encouraging earlier detection and intervention.

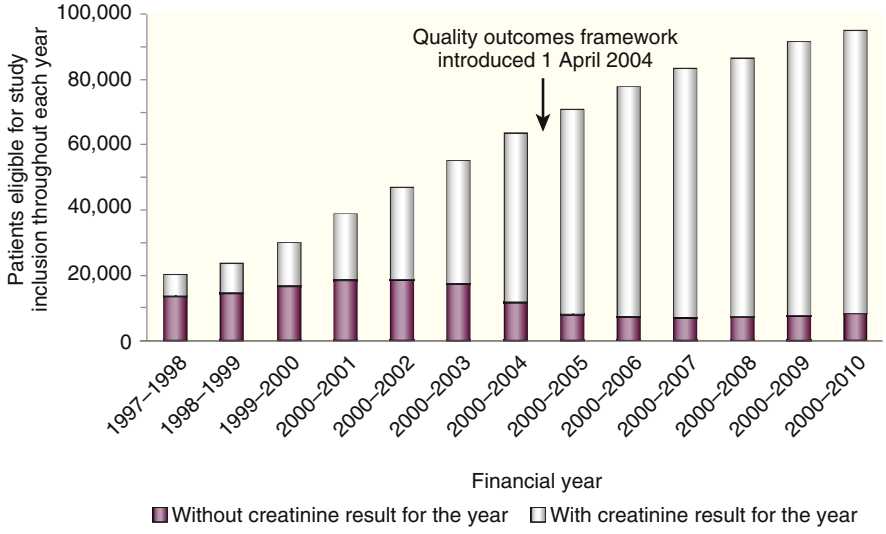

Figure 2 | UK example: completeness of serum creatinine recording among 219,145 patients with diabetes aged ≥65 years for each full financial year. Patients were eligible for inclusion into the study if they were registered within Clinical Practice Research Datalink practices fulfilling quality criteria of reporting, aged 65 years or more, and had codes consistent with having diabetes.[30]

Knowledge and awareness of the diagnostic criteria for AKI among clinicians in primary and secondary care is changing rapidly but has been limited until recently. While it has been poorly investigated to date, it is likely that coding for AKI (or acute renal failure) in primary care reflects the situation in secondary care: coding captures a fraction of AKI cases defined by biochemical criteria and is particularly likely to miss less severe cases.[21] Accuracy of coding of abnormal renal function also relies on the physician being able to distinguish between AKI and CKD.

The need for clinical recognition can be removed by applying AKI staging criteria to biochemical data.[22] However, AKI algorithms may misclassify worsening CKD as AKI; in a hospital setting, 14% of those identified had CKD rather than AKI.[21] As the AKI algorithm was developed for use with inpatient laboratory data, it should be used with caution in community data. Compared with a hospital setting, serum creatinine measurements in the community are likely to be separated by longer intervals. This increases the likelihood of misclassifying a gradual decline in renal function as AKI. In addition, it is not possible to apply AKI staging criteria to a single serum creatinine measure without knowledge of a previous and reliable baseline. Patients with available baseline renal function are unlikely to be representative of the general population; they are likely to be either people with chronic conditions that have prompted testing or those who engage more with health care. Previously healthy individuals are less likely to have recent baseline serum creatinine results available.

What other data items are needed for confounding control and are these items captured in the database?

Similar to what was discussed for renal disease, guideline changes and performance incentives may result in changes to how information on other key variables is recorded over time. For example, a financially incentivized change to the recommended management of patients diagnosed with depression led general practitioners to use fewer depression diagnosis codes and use nonspecific symptom codes instead.[23] If such issues are not considered, narrow definitions may result in incomplete capture of key variables.

Certain items may not be recorded at all as they are outside the remit of care. For example, although in the UK most prescriptions are issued by primary care physicians, key drugs of interest to renal physicians (such as erythropoietin, certain immunosuppressant drugs, and biologics [e.g., rituximab]) are commissioned and prescribed in secondary care, which means that primary care data on these drugs will be incomplete. Another example is nonsteroidal anti-inflammatory drug use, which cannot be captured accurately in settings where these drugs are freely available over the counter.

Finally, when extracting dates of diagnoses of certain diseases, it is important to consider timing relative to registration with a doctor. Disease symptoms may lead some patients to join with a new practice, biasing observed incidence rates upward for the time shortly after registration. Preexisting comorbidities may also be entered without distinction from new diagnoses during early patient visits in which previous medical history is established. Lewis *et al.*[24] found that this increased incidence rate following new patient registration returned to baseline within 6 months for most acute conditions and within a year for most chronic conditions. This means that analyses of incident disease need to have a start time of observation at least 6 months after registration, and a year is commonly used.

Can potential biases when using EHRs be mitigated for renal analyses?

In this section, the 2 most common observational study designs (case-control, cohort) will be used as a device to discuss common biases in renal EHRs research, with eGFR as a baseline variable and AKI as an outcome variable (Table 2).

Minimizing confounding. Testing for eGFR and proteinuria in the general population is often influenced by clinical characteristics. As long as the researcher investigates these characteristics and understands testing incentives over time, it is possible to adapt the study design and analyses for these confounding factors. If all test results are used, adjustment for health status is needed to avoid residual confounding, but aspects of health status may be poorly captured in EHRs. Alternatively, restriction of the epidemiologic study to a cohort with higher cardiovascular risk who are expected to be routinely tested ensures that the indication for the baseline renal function test is independent from its test result.

Understanding how clinicians prescribe is key to interpreting findings and spotting potential sources of residual confounding. For example, certain medications may be prescribed more or less commonly in patients with CKD, thus presence or absence of a drug prescription may be a surrogate marker of underlying kidney function. Angiotensin-converting enzyme inhibitors may, for example, be prescribed more commonly in people with proteinuria on dipstick testing. However, if proteinuria is recorded less completely than angiotensin-converting enzyme inhibitor prescriptions, then analyses using recorded data will be subject to residual confounding by uncoded proteinuria.

Avoiding exposure misclassification and immortal time bias. In traditional epidemiological studies of the effect of CKD on outcomes, the definition of CKD is often based on a single measurement of serum creatinine at baseline when the study participant is in steady-state. However, identifying CKD based on a single estimate of GFR in clinical records will tend to overestimate CKD prevalence, because a single impaired GFR may also be caused by AKI or because the test may have been prompted by ill health.[25] Therefore, the clinical classification of CKD requires that impaired renal function persists for at least 3 months for CKD diagnosis. However, if 2 results are required to define time-updated CKD status in epidemiological studies, the time between the first and second results must be handled with care to avoid immortal time bias. Patients have to be alive in order to attend clinic and have the

Table 2 | Summary of potential biases that may arise in an example study of the association between baseline eGFR and AKI

	Potential biases				
	Immortal time bias	**Selection bias**	**Competing risk**	**Outcome misclassification**	**Reverse causality**
Cohort study	Allocating time that does not have outcome events systematically to 1 exposure group.	Differential completeness of follow-up, because specific subgroups do not remain in the study population or attend facility that collects outcome data.	Censoring of follow-up time differentially by exposure status.	Outcome is not systematically identified and may be differential or nondifferential.	Outcome data affect exposure data differentially.
Example	Only coding patients with 2 measurements of eGFR <60 as having CKD and starting follow-up at first eGFR measurement. This assumes that patients do not have AKI after a single eGFR measurement.	Patients with AKI who do not require chronic dialysis may not be recorded in renal registry, or cannot afford hospitalization or dialysis, and therefore are not recorded in the registry.	When exploring AKI rates in the very elderly consider that study participants with more advanced stages of CKD may die before having AKI.	Coded AKI on discharge is used as an outcome variable. Differential misclassification occurs if doctors may be more aware of AKI if patient has underlying known kidney disease.	AKI is defined by coded admission; however, AKI occurred earlier and the primary care physician took blood to measure eGFR during the acute illness, and then referred patient to hospital.
Suggested mitigation for example	Use last carried forward method (Figure 3), or define CKD status at the second time-point comparing with people who have a similar time of follow-up between the first and second measurements.	Restrict study population to those who can afford health care and use laboratory values at admission and compare with previous values >1 week before admission.	Competing risk analysis.	Carry out validation studies of AKI coding to explore this, if coding is nondifferential (not dependent on exposure status) study findings are likely to be an underestimate of the true underlying association.	Remove eGFR values in week before hospital admission.
Case control study	Akin to selection bias in case-control study setting if sampling of control subjects is restricted to living individuals.	(i) Control subjects are not drawn from the population that case subjects are derived from, or (ii) AKI cases are not representative of the population who have AKI.	Akin to selection bias in the case-control setting.	Control subjects may contain some undiagnosed cases, and/or some cases may not have AKI.	Case status influences exposure data differentially from those in controls.
Example	(i) AKI cases are identified in a hospital database and compared with control subjects sampled from primary care, some of whom cannot afford to attend the hospital that was used to identify cases. (ii) Only including AKI cases recorded in registry, but the registry has poor coverage and is more likely to capture cases with CKD.			See example and corresponding mitigation above.	See example and corresponding mitigation above.
Suggested mitigation for example	(i) Source control subjects from a population of hospital attenders. (ii) Use alternative sources of identifying AKI.				

Potential ways that bias may be mitigated are given but are not exhaustive. Potential confounding factors are not listed/discussed in this table but must also be addressed. AKI, acute kidney injury; CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate.

second creatinine test after 3 months, and therefore CKD should not be defined at the time of the first eGFR measurement. A previous study compared people who had both a test and a follow-up clinic visit with people who did not attend the follow-up clinic visit and found that this may introduce an artificial survival advantage for those who attended follow-up of at least 20%.[26] Alternatively, the comparison should be between people who have CKD as defined by the second eGFR result versus individuals with at least the same length of follow-up in the database. The "last-carried-forward" time-updated method used by James *et al.*[27] may be more suitable for studies with follow-up time long enough for CKD progression to be an issue. In this approach, CKD stage is defined at any given time using the GFR estimate produced by the single most recent creatinine result (Figure 3). This method allows updating of the patient's status as CKD progresses. Although creatinine fluctuation will result in misclassification of CKD stage, the patient's status will be updated at the next test result, minimizing misclassified person-time.

Avoiding selection bias and outcome misclassification. The best way to avoid selection bias is to restrict the study

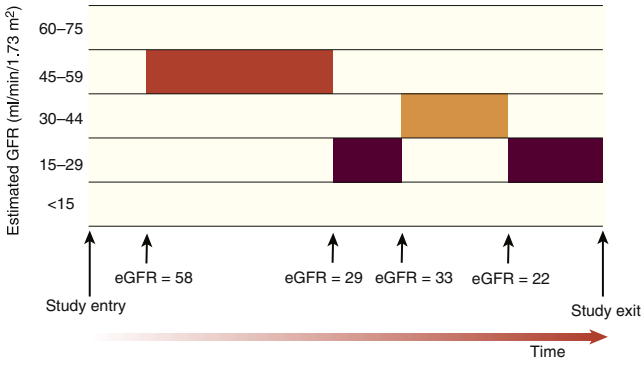

Figure 3 | The last-carried-forward method for establishing eGFR over time according to a patient's eGFR category. Shaded bars correspond to the time a patient is thought to have a particular estimated glomerular filtration rate (eGFR) category, with arrows representing study entry and times of recorded eGFR results in the health record.[30]

population to those for whom it is true that if the outcome occurs, it will be recorded in the participant's health record, and that control subjects in a case-control study are drawn from the same population as the case subjects. If the study outcome is starting dialysis, then the start date of dialysis needs to be clearly defined to avoid misclassification. For example, it is important to clarify whether for those on peritoneal dialysis the start date is the date of insertion of a peritoneal dialysis catheter or the start of peritoneal dialysis training. A complicating factor is that the EHRs may span a period of time over which identification (for example, see Figures 1 and 2) and recording of CKD and AKI change,[28] affecting how those records may be interpreted. Features of patients with AKI recorded at hospital admission may change over time because AKI is an increasingly recognised complication in hospitalized patients.[29] Exploration of the time-dependent nature of these issues may require sensitivity analyses in subsets of the cohort stratified by calendar period. Physicians may be actively looking for AKI in at-risk populations. This can be explored by validation studies or addressed by sensitivity analyses restricted to at-risk populations.

Conclusions

Using routine data from EHRs provides potential for large cohorts to be derived with detailed longitudinal records to facilitate a wide range of kidney research. However, challenges exist in terms of availability and validity of measures of renal function. Understanding what influences testing and recording of data is vital when planning research using routinely collected data. Reporting of observational studies using EHRs should adhere to the Reporting of Studies Conducted Using Observational Routinely-Collected Health Data (RECORD) guidelines.[6]

DISCLOSURE

All the authors declared no competing interests.

ACKNOWLEDGMENTS

HIM was funded by Kidney Research UK (ST2/2011). SLT was funded by a Career Development Fellowship from the National Institute for Health Research [grant number CDF 2010-03-32]. LAT is funded by a Wellcome Trust intermediate clinical fellowship (101143/Z/13/Z).

The study funders had no role in the preparation, review, nor approval of the manuscript, nor the decision to submit it for publication. The views expressed in this publication are those of the authors and not necessarily those of the UK National Health Service, the National Institute for Health Research, or the Department of Health.

REFERENCES

1. Mehta RL, Cerdá J, Burdmann EA, et al. International Society of Nephrology's 0by25 initiative for acute kidney injury (zero preventable deaths by 2025): a human rights case for nephrology. *Lancet*. 2011;385:2616–2643.
2. Department of Economic and Social Affairs Population Division. World Population Ageing 2013. United Nations, 2013. Available at: http://www.un.org/en/development/desa/population/publications/pdf/ageing/WorldPopulationAgeing2013.pdf. Accessed February 3, 2016.
3. National Institute for Health and Care Excellence. Acute kidney injury: prevention, detection and management 2013. Available at: https://www.nice.org.uk/guidance/cg169. Accessed February 3, 2016.
4. McDonald HI, Thomas SL, Millett ER, Nitsch D. CKD and the risk of acute, community-acquired infections among older people with diabetes mellitus: a retrospective cohort study using electronic health records. *Am J Kidney Dis*. 2015;66:60–68.
5. de Lusignan S, Nitsch D, Belsey J, et al. Disparities in testing for renal function in UK primary care: cross-sectional study. *Fam Pract*. 2011;28:638–646.
6. Benchimol EI, Smeeth L, Guttmann A, et al, for the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Med*. 2015;12:e1001885.
7. James MT, Pannu N. Methodological considerations for observational studies of acute kidney injury using existing data sources. *J Nephrol*. 2009;22:295–305.
8. Moosa MR, Kidd M. The dangers of rationing dialysis treatment: the dilemma facing a developing country. *Kidney Int*. 2006;70:1107–1114.
9. White SL, Chadban SJ, Jan S, Chapman JR, Cass A. How can we achieve global equity in provision of renal replacement therapy? *Bull World Health Organ*. 2008;86:229–237.
10. Hemmelgarn BR, James MT, Manns BJ, et al, for the Alberta Kidney Disease Network. Rates of treated and untreated kidney failure in older vs younger adults. *JAMA*. 2012;307:2507–2515.
11. Garg AX, Mamdani M, Juurlink DN, et al, for the Network of Eastern Ontario Medical Laboratories. Identifying individuals with a reduced GFR using ambulatory laboratory database surveillance. *J Am Soc Nephrol*. 2005;16:1433–1439.
12. National Institute for Health and Care Excellence. Chronic kidney disease in adults: assessment and management 2014. Available at: https://www.nice.org.uk/guidance/cg182. Accessed February 3, 2016.
13. Vlasschaert ME, Bejaimal SA, Hackam DG, et al. Validity of administrative database coding for kidney disease: a systematic review. *Am J Kidney Dis*. 2011;57:29–43.
14. van Walraven C, Austin PC, Manuel D, et al. The usefulness of administrative databases for identifying disease cohorts is increased with a multivariate model. *J Clin Epidemiol*. 2010;63:1332–1341.
15. Caplin B, Wheeler DC, Nitsch D, Hull S. The National Chronic Kidney Disease Audit Pilot Report. 2015. Available at: http://www.hqip.org.uk/public/cms/253/625/24/96/Kidney%20-%20National-CKD-Audit-Pilot-Report-2014.pdf?realName=lFyhnE.pdf. Accessed February 3, 2016.
16. Anandarajah S, Tai T, de Lusignan S, et al. The validity of searching routinely collected general practice computer data to identify patients with chronic kidney disease (CKD): a manual review of 500 medical records. *Nephrol Dial Transplant*. 2005;20:2089–2096.
17. Denburg MR, Haynes K, Shults J, et al. Validation of The Health Improvement Network (THIN) database for epidemiologic studies of chronic kidney disease. *Pharmacoepidemiol Drug Saf*. 2011;20:1138–1149.

18. Poh N, McGovern A, De Lusignan S. Improving the measurement of longitudinal change in renal function: automated detection of changes in laboratory creatinine assay. *J Innov Health Inform*. 2015;22:292–301.
19. Hemmelgarn BR, Manns BJ, Lloyd A, et al, for the Alberta Kidney Disease Network. Relation between kidney function, proteinuria, and adverse outcomes. *JAMA*. 2010;303:423–429.
20. Bellomo R, Ronco C, Kellum JA, et al, for the the Acute Dialysis Quality Initiative Workgroup. Acute renal failure—definition, outcome measures, animal models, fluid therapy and information technology needs: the Second International Consensus Conference of the Acute Dialysis Quality Initiative (ADQI) Group. *Crit Care*. 2004;8:R204–212.
21. Sawhney S, Marks A, Ali T, et al. Maximising acute kidney injury alerts – a cross-sectional comparison with the clinical diagnosis. *PLoS One*. 2015;10:e0131909.
22. Association for Clinical Biochemistry and Laboratory Medicine. Algorithm for generating E-Alerts for Acute Kidney Injury based on serum creatinine changes with time 2013. Available at: http://www.acb.org.uk/docs/appendix-a-algorithm. Accessed February 3, 2016.
23. Kendrick T, Stuart B, Newell C, et al. Changes in rates of recorded depression in English primary care 2003–2013: time trend analyses of effects of the economic recession, and the GP contract quality outcomes framework (QOF). *J Affect Disord*. 2015;180:68–78.
24. Lewis JD, Bilker WB, Weinstein RB, Strom BL. The relationship between time since registration and measured incidence rates in the General Practice Research Database. *Pharmacoepidemiol Drug Saf*. 2005;14:443–451.
25. de Lusignan S, Tomson C, Harris K, et al. Creatinine fluctuation has a greater effect than the formula to estimate glomerular filtration rate on the prevalence of chronic kidney disease. *Nephron Clin Pract*. 2011;117:c213–224.
26. 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.
27. James MT, Quan H, Tonelli M, et al, for the Alberta Kidney Disease Network. CKD and risk of hospitalization and death with pneumonia. *Am J Kidney Dis*. 2009;54:24–32.
28. Waikar SS, Curhan GC, Wald R, et al. Declining mortality in patients with acute renal failure, 1988 to 2002. *J Am Soc Nephrol*. 2006;17:1143–1150.
29. Kolhe NV, Muirhead AW, Wilkes SR, et al. The epidemiology of hospitalised acute kidney injury not requiring dialysis in England from 1998 to 2013: retrospective analysis of hospital episode statistics. *Int J Clin Pract*. 2016;70:330–339.
30. McDonald HI. *The Epidemiology of Infections among Older People with Diabetes Mellitus and Chronic Kidney Disease* [dissertation]. London, UK: London School of Hygiene and Tropical Medicine; 2015.