



RESEARCH ARTICLE

Statistical models for the deterioration of kidney function in a primary care population: A retrospective database analysis [version 1; peer review: 1 approved with reservations]

Jason L Oke ¹, Benjamin G Feakins ¹, Iryna Schlackow²,
 Borislava Mihaylova ^{2,3}, Claire Simons², Chris A O'Callaghan ⁴,
 Daniel S Lasserson⁵, F D Richard Hobbs¹, Richard J Stevens ¹, Rafael Perera ¹

¹Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, OX2 6GG, UK
²Nuffield Department of Population Health, University of Oxford, Oxford, OX3 7LF, UK
³Centre for Primary Care and Public Health, Queen Mary, University of London, London, E1 2AB, UK
⁴Nuffield Department of Medicine, University of Oxford, Oxford, OX3 7LF, UK
⁵Institute of Applied Health Research, University of Birmingham, Birmingham, B15 2TT, UK

v1 **First published:** 10 Sep 2019, 8:1618 (<https://doi.org/10.12688/f1000research.20229.1>)
Latest published: 10 Sep 2019, 8:1618 (<https://doi.org/10.12688/f1000research.20229.1>)

Abstract

Background: Evidence for kidney function monitoring intervals in primary care is weak, and based mainly on expert opinion. In the absence of trials of monitoring strategies, an approach combining a model for the natural history of kidney function over time combined with a cost-effectiveness analysis offers the most feasible approach for comparing the effects of monitoring under a variety of policies. This study aimed to create a model for kidney disease progression using routinely collected measures of kidney function.

Methods: This is an open cohort study of patients aged ≥18 years, registered at 643 UK general practices contributing to the Clinical Practice Research Datalink between 1 April 2005 and 31 March 2014. At study entry, no patients were kidney transplant donors or recipients, pregnant or on dialysis. Hidden Markov models for estimated glomerular filtration rate (eGFR) stage progression were fitted to four patient cohorts defined by baseline albuminuria stage; adjusted for sex, history of heart failure, cancer, hypertension and diabetes, annually updated for age.

Results: Of 1,973,068 patients, 1,921,949 had no recorded urine albumin at baseline, 37,947 had normoalbuminuria (<3mg/mmol), 10,248 had microalbuminuria (3–30mg/mmol), and 2,924 had macroalbuminuria (>30mg/mmol). Estimated annual transition probabilities were 0.75–1.3%, 1.5–2.5%, 3.4–5.4% and 3.1–11.9% for each cohort, respectively. Misclassification of eGFR stage was estimated to occur in 12.1% (95%CI: 11.9–12.2%) to 14.7% (95%CI: 14.1–15.3%) of tests. Male gender, cancer, heart failure and age were independently associated with declining renal function, whereas the impact of raised blood pressure and glucose on renal function was entirely predicted by albuminuria.

Conclusions: True kidney function deteriorates slowly over time, declining

Open Peer Review

Reviewer Status ?

	Invited Reviewers
	1
version 1 published 10 Sep 2019	? report

1 **Dorothea Nitsch**, London School of Hygiene and Tropical Medicine (LSHTM), London, UK

Any reports and responses or comments on the article can be found at the end of the article.

more sharply with elevated urine albumin, increasing age, heart failure, cancer and male gender. Consecutive eGFR measurements should be interpreted with caution as observed improvement or deterioration may be due to misclassification.

Keywords

Kidney Function Decline, Chronic Kidney Disease (CKD), Estimated Glomerular Filtration Rate (eGFR), Proteinuria, Hidden Markov Model (HMM), Primary Care, Clinical Practice Research Datalink (CPRD)

Corresponding author: Benjamin G Feakins (benjamin.feakins@phc.ox.ac.uk)

Author roles: **Oke JL:** Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Feakins BG:** Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Schlackow I:** Investigation, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Mihaylova B:** Investigation, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Simons C:** Investigation, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **O’Callaghan CA:** Funding Acquisition, Investigation, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Lasserson DS:** Funding Acquisition, Investigation, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Hobbs FDR:** Funding Acquisition, Investigation, Methodology, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Stevens RJ:** Conceptualization, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Perera R:** Conceptualization, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This research was funded by the National Institute for Health Research (NIHR) Programme Grants for Applied Research programme [Reference: RP-PG-1210-12003]. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. RP, FDRH and BM receive funding from the NIHR Oxford Biomedical Research Council (BRC). RP and FDRH receive funding from the NIHR Collaboration for Leadership in Health Research and Care (CLARHC) Oxford and the NIHR Oxford Medtech and In-Vitro Diagnostics Co-operative (MIC). RP receives funding from the NIHR Health Protection Research Unit Gastrointestinal Infections Group. FDRH receives funding from the NIHR School for Primary Care Research.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2019 Oke JL *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Oke JL, Feakins BG, Schlackow I *et al.* **Statistical models for the deterioration of kidney function in a primary care population: A retrospective database analysis [version 1; peer review: 1 approved with reservations]** F1000Research 2019, 8:1618 (<https://doi.org/10.12688/f1000research.20229.1>)

First published: 10 Sep 2019, 8:1618 (<https://doi.org/10.12688/f1000research.20229.1>)

Introduction

The National Institute for Health and Care Excellence recommend monitoring kidney function using estimated glomerular filtration rate (eGFR) in people with, or at risk of, chronic kidney disease (CKD)¹. The guideline suggests increasing the intensity of monitoring according to the current level of eGFR and albumin-creatinine ratio, stating that monitoring should be tailored according to i) the underlying cause of CKD and ii) past patterns of eGFR and albumin-creatinine ratio, comorbidities, changes to treatments such as renin-angiotensin-aldosterone system antagonists, inter-current illness and whether the patient has chosen conservative management of CKD. One of the objectives of monitoring eGFR is to detect progression of CKD, which could precede end-stage renal disease (ESRD). ESRD is associated with substantial morbidity and mortality, with cardiovascular disease mortality rates 10 to 30 times higher in patients on dialysis than in the general population². Yet, kidney function declines slowly with age and ESRD is rare, even for people with moderately impaired renal function (eGFR 30–59 ml/min/1.73m²). In a study of 58,000 people with CKD stage 3 who were followed for 10 years, the cumulative incidence was 40 per 1,000 people³. It follows that recommendations to monitor everyone annually or more frequently in a community setting for progressive kidney function loss will have a poor yield. Furthermore, as eGFR is a noisy measurement, with a within-person coefficient of variation estimated to be approximately 5.5%⁴, it is likely two consecutive eGFR measurements may appear to indicate declining renal function when underlying renal function is stable (false positive), or stable renal function when underlying renal function has deteriorated (false negative). Finally, it is arguable as to whether there are any actions that can be taken to halt the deterioration of renal function if progressive CKD is found, as there is currently very little evidence that “catching” CKD early produces any benefit⁵.

There have been no trials of screening or monitoring for CKD⁶ and recommendations for how frequently monitoring should take place are based on expert opinion. In the absence of trials, an approach combining a model for the natural history of kidney function over time combined with a cost-effectiveness analysis offers the most feasible approach for comparing the effects of monitoring under a variety of policies. The aim of this study was to create a model for kidney disease progression using routine measures of kidney function. Our approach simultaneously estimates the true rate of kidney function loss and the probability of misclassification that inevitably occurs from using eGFR. Our study is conducted in a general primary care population and our results will be useful in guiding future recommendations for the timing of monitoring eGFR in primary care.

Methods

Ethical statement

The protocol for this research was approved by the Independent Scientific Advisory Committee of the Medicines and Healthcare Products Regulatory Agency (protocol number 14_150R). Ethical approval for observational research using the Clinical Practice Research Datalink with approval from the

Independent Scientific Advisory Committee has been granted by a National Research Ethics Service committee (Trent Multi Research Ethics Committee, REC reference number 05/MRE04/87).

Source and selection of participants

We used the UK Clinical Practice Research Practice Datalink (CPRD)⁷ to construct an open cohort of adults (≥ 18 years of age) registered at practices deemed to have “acceptable” patient records (termed “up-to-standard” in CPRD). We included patient records starting from 1 April 2005, post-dating the publication of the Kidney Disease Outcomes Quality Initiative (KDOQI) guidelines for the classification of CKD in 2002⁸ and the introduction of Quality and Outcomes Framework targets in UK primary care in 2004. The study end date was 31 March 2014. Eligible patients had to be registered with their practice for a minimum of 12 months before study entry to ensure adequate recording of baseline covariates. We excluded patients who, in the 12 months before study entry, were pregnant, were receiving dialysis, or were living kidney donors or recipients. Follow-up ended at the study end date, unless preceded by the date of death, transfer out of CPRD, the last available linked data, or (where applicable) pregnancy, renal transplantation/donation, or dialysis.

Statistical analysis

To model decline in kidney function, hidden Markov models (HMMs)^{9–13} were fitted to four patient cohorts defined by baseline albuminuria stage: 1) no albuminuria measurement (unmeasured), 2) normoalbuminuria (< 3 mg/mmol), 3) microalbuminuria (3–30 mg/mmol), and 4) macroalbuminuria (> 30 mg/mmol). Models were adjusted for sex, heart failure, cancer, hypertension and diabetes, and annually updated age.

The HMMs comprised two components, a multi-state model governing the ‘true’ underlying progression of CKD, and a second model for the probability of misclassification to allow for the variability in eGFR. The underlying model for CKD was parametrised as uni-directional, in which true kidney function could only deteriorate over time (no spontaneous improvement). The outcome was eGFR stage based on the criteria used for the diagnosis of CKD, i.e. G1–G5. We combined stages G1 and G2 for the purposes of improving model fit. Death from any cause was assumed to be an absorbing state. A representation of the HMMs is depicted in [Figure 1](#).

The HMMs were specified so that it was possible for misclassification to occur in neighbouring eGFR categories. Hence, for a person with true GFR > 60 ml/min/1.73m² we specified the model so that a single measurement of eGFR could fall within a G3a or G3b category due to measurement error and biological variation, but not G4 or G5. For a person with true eGFR in stage G3b, a single measurement of eGFR could be misclassified as either G1/2, G3a, G4 or G5. Death was the only state assumed to be always classified correctly.

To assess model fit, we used a split-sample approach. Although this is a weak procedure for low-variance methods, such as the Cox proportional hazards model or logistic regression,

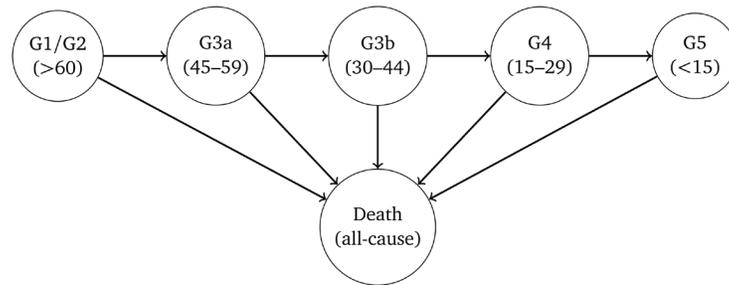


Figure 1. Representation of the model for the deterioration of kidney function over time. Arrows indicate permitted (instantaneous) transitions. The numbers in brackets depict the estimated glomerular filtration rate ranges (in ml/min/1.73²) associated with each stage.

it is useful for a model that can be over-parametrised or exhibit convergence issues (such as a HMM). We split the data using pseudo-random numbers into equal size training and testing data sets. The model was fit in the training data set and then used to predict trajectories of eGFR for patients in the testing data set, based on their measurement times and covariates. Calibration plots were used to compare the predicted and observed proportion of tests falling within each eGFR category over time. Annual transition rates for kidney function loss and death from any cause were estimated from the model, along with the misclassification probabilities and transition rate multipliers for age, sex, heart failure and cancer, and presented as state model diagrams. The models were used to estimate the probability of progression to a higher stage within six, 12 or 36 months, along with the probability that an eGFR test taken at that time would detect the change (true positive), and the probability that a change in eGFR stage would occur in a person in whom true kidney function had not changed (false positive), for all cohorts for baseline stages G3a and G3b; see Supplementary Tables S18–21 (*Extended data*)¹⁴.

Finally, we estimated global misclassification probabilities for the four cohorts using the Viterbi algorithm¹⁵ to find the underlying sequence of true eGFR stages with the highest probability given the observed sequence. Assuming the state predicted by the model was the truth, we calculated the proportion of times the observed state was a lower stage than predicted (under-grading) and the proportion of times the observed was a higher stage than predicted (over-grading), and then added these together to calculate the total number of misclassified tests across cohorts.

All analyses were performed in R version 3.6.1 (“Action of the Toes”)¹⁶, with HMMs fit using version 1.6.7 of the *msm* package¹⁷. Scripts used in these analyses are available (see *Software availability*)¹⁸.

Results

The initial data set comprised 3,338,526 patients. A total of 1,365,458 patients whose records contained fewer than three eGFR tests were excluded, leaving 1,973,068 patients eligible for analysis: 1,921,949 without a urine albumin test on record,

37,947 with normoalbuminuria (<3 mg/mmol), 10,248 with microalbuminuria (3–30 mg/mmol), and 2,924 with macroalbuminuria (>30 mg/mmol). Each of the four cohorts were split into two halves and nominated as training and testing data sets. Due to the computational demands of the statistical method used, we randomly selected a sub-cohort of 50,000 patients to fit the model in the cohort without a urine albumin test on record. Summary statistics of patient characteristics from the four cohorts are presented in [Table 1](#).

Six state continuous time HMMs adjusted for sex, heart failure, cancer, hypertension and diabetes, and annually updated age were fit on the four training data sets. Hypertension and diabetes were subsequently removed from the models as they were unable to predict eGFR stage progression or death. All models converged to their respective maximum likelihood estimates, with positive definitive Hessian matrices permitting confidence interval estimation for all parameters. Intensity, transition and misclassification matrices for these models are given in Supplementary Tables S2–13 (*Extended data*)¹⁴.

[Figure 2](#) shows the annual transition and misclassification probabilities for a woman, aged 60, without heart failure or a previous diagnosis of cancer and with no urine albumin test on record. The figure shows that if kidney function is normal (G1/G2) then the probability of her true kidney function deteriorating to stage G3a in one year is estimated to be 1.1%. The probability that a single eGFR test will be misclassified as G3a is 2.9%, while the probability that it will correspond to her true stage is 97.1%. The probability that this woman dies within a year is estimated to be 0.7%. The probability that her kidney function remains in this category is 98.2%. If the woman is one year older then transition probabilities should be multiplied by 1.08 for kidney function and 1.09 for death. For example, the annual transition probability from stage G3b, is 1.0% for a 60 year old woman, but $1.0 \times 1.08^{10} = 2.16\%$ for a 70 year old woman and $\times 1.08^{20} = 4.66\%$ for woman who is 80 years old. Multipliers in which the confidence interval overlapped “no effect” are set to 1.00.

[Figure 3](#) represents annual transitions for a woman with the same characteristics, but who has had her urine albumin

Table 1. Patient characteristics at baseline, by albuminuria stage.

Variable	Category	Albuminuria Stage, Number (%)			
		Unmeasured	Normoalbuminuria	Microalbuminuria	Macroalbuminuria
	Total	1,921,949 (100.0%)	37,947 (100.0%)	10,248 (100.0%)	2,924 (100.0%)
Gender	Female	1,058,400 (55.1%)	18,312 (48.3%)	4,749 (46.3%)	1,352 (46.2%)
	Male	863,549 (44.9%)	19,635 (51.7%)	5,499 (53.7%)	1,572 (53.8%)
Age (years)	18–39	254,037 (13.2%)	2,701 (7.1%)	502 (4.9%)	267 (9.1%)
	40–49	324,362 (16.9%)	4,811 (12.7%)	928 (9.1%)	352 (12.0%)
	50–59	419,561 (21.8%)	7,354 (19.4%)	1,547 (15.1%)	514 (17.6%)
	60–69	421,704 (21.9%)	9,930 (26.2%)	2,191 (21.4%)	638 (21.8%)
	70–79	321,522 (16.7%)	8,610 (22.7%)	2,575 (25.1%)	602 (20.6%)
	80–89	154,881 (8.1%)	3,948 (10.4%)	2,006 (19.6%)	444 (15.2%)
	90+	25,882 (1.3%)	593 (1.6%)	499 (4.9%)	107 (3.7%)
Ethnicity	Missing	1,133,893 (59.0%)	18,469 (48.7%)	4,950 (48.3%)	1,821 (62.3%)
	White	461,796 (24.0%)	10,077 (26.6%)	2,465 (24.1%)	515 (17.6%)
	Black	28,480 (1.5%)	1,325 (3.5%)	512 (5.0%)	106 (3.6%)
	Asian	12,314 (0.6%)	728 (1.9%)	229 (2.2%)	47 (1.6%)
	Mixed	276,750 (14.4%)	6,999 (18.4%)	1,958 (19.1%)	422 (14.4%)
	Other	8,716 (0.5%)	349 (0.9%)	134 (1.3%)	13 (0.4%)
eGFR (ml/min/1.73m ²)	>60	1,524,003 (79.3%)	27,753 (73.1%)	6,114 (59.7%)	1,628 (55.7%)
	45–59	295,312 (15.4%)	6,850 (18.1%)	2,147 (21.0%)	569 (19.5%)
	30–44	85,303 (4.4%)	2,724 (7.2%)	1,440 (14.1%)	453 (15.5%)
	15–29	16,091 (0.8%)	591 (1.6%)	518 (5.1%)	247 (8.4%)
	<15	1,240 (0.1%)	29 (0.1%)	29 (0.3%)	27 (0.9%)
CKD Read Code	None	1,911,565 (99.5%)	37,521 (98.9%)	10,044 (98.0%)	2,870 (98.2%)
	G1/2	2,660 (0.1%)	122 (0.3%)	23 (0.2%)	5 (0.2%)
	G3	7,347 (0.4%)	282 (0.7%)	148 (1.4%)	31 (1.1%)
	G4	457 (0.0%)	21 (0.1%)	34 (0.3%)	17 (0.6%)
	G5	87 (0.0%)	1 (0.0%)	0 (0.0%)	2 (0.1%)
Cancer	No	1,884,014 (98.0%)	37,698 (99.3%)	10,206 (99.6%)	2,912 (99.6%)
	Yes	37,935 (2.0%)	249 (0.7%)	42 (0.4%)	12 (0.4%)
Chronic Renal Disease	No	1,919,946 (99.9%)	37,928 (99.9%)	10,230 (99.8%)	2,914 (99.7%)
	Yes	2,003 (0.1%)	19 (0.1%)	18 (0.2%)	10 (0.3%)
Diabetes	No	1,866,051 (97.1%)	35,850 (94.5%)	9,660 (94.3%)	2,810 (96.1%)
	Yes	55,898 (2.9%)	2,097 (5.5%)	588 (5.7%)	114 (3.9%)
Heart Failure	No	1,905,724 (99.2%)	37,778 (99.6%)	10,209 (99.6%)	2,914 (99.7%)
	Yes	16,225 (0.8%)	169 (0.4%)	39 (0.4%)	10 (0.3%)
Hypertension	No	1,512,801 (78.7%)	34,353 (90.5%)	9,541 (93.1%)	2,749 (94.0%)
	Yes	409,148 (21.3%)	3,594 (9.5%)	707 (6.9%)	175 (6.0%)
Ischaemic Heart Disease	No	1,841,610 (95.8%)	37,275 (98.2%)	10,122 (98.8%)	2,897 (99.1%)
	Yes	80,339 (4.2%)	672 (1.8%)	126 (1.2%)	27 (0.9%)
Peripheral Vascular Disease	No	1,895,750 (98.6%)	37,766 (99.5%)	10,213 (99.7%)	2,912 (99.6%)
	Yes	26,199 (1.4%)	181 (0.5%)	35 (0.3%)	12 (0.4%)
Stroke or TIA	No	1,890,775 (98.4%)	37,695 (99.3%)	10,189 (99.4%)	2,905 (99.4%)
	Yes	31,174 (1.6%)	252 (0.7%)	59 (0.6%)	19 (0.6%)

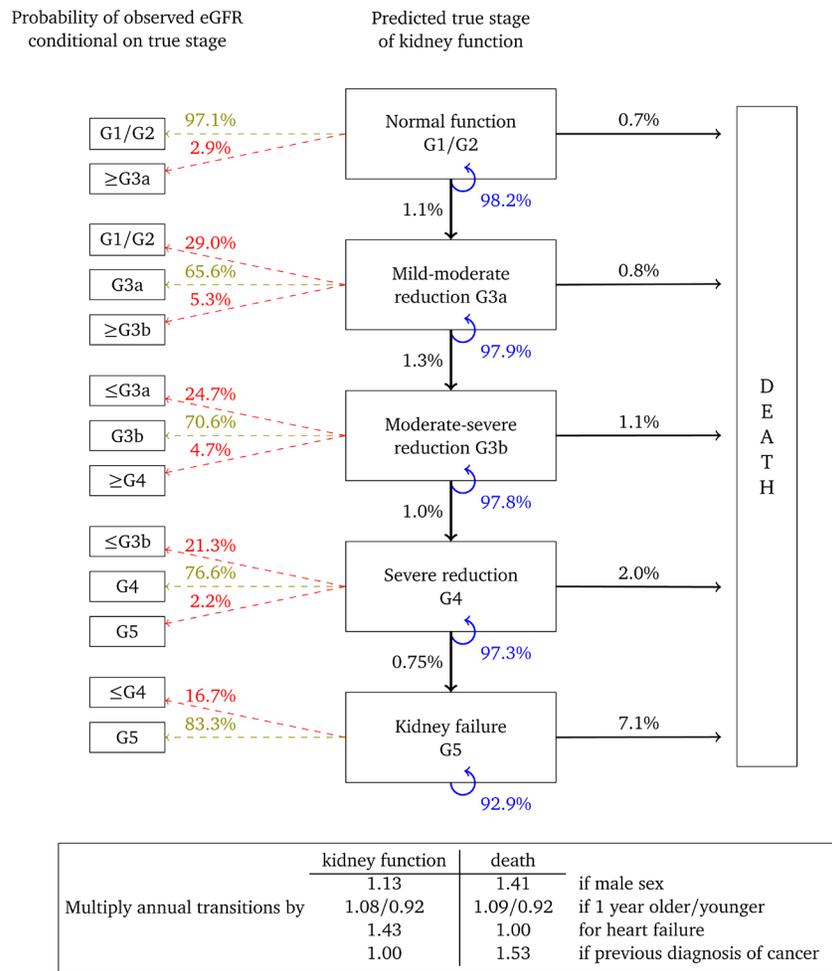


Figure 2. Annual transition model diagram for patients with unmeasured urine albumin at baseline. Probabilities are based on a woman aged 60, without heart failure or a previous diagnosis of cancer.

tested and found to be in the normoalbuminuric range. Corresponding annual transition probabilities for kidney function are nearly twice that of an equivalent woman without a urine albumin test on record. Respective transition rates to death from each stage are also higher, illustrating that this cohort represents women in poorer health. Misclassification probabilities and transition probability multipliers are broadly similar to Figure 2.

Figure 4 and Figure 5 show results for women with micro- and macroalbuminuria, respectively. Kidney function transition probabilities are higher, as are annual transition probabilities for death. Fewer transition multipliers are significant for these cohorts but this probably reflects the smaller cohort sizes and correspondingly reduced statistical power.

Table 2 shows the results from applying the Viterbi algorithm to the four cohorts. Under-grading of eGFR stage occurs more often than over-grading in all cohorts but over-grading tends to increase for cohorts having urine albumin tests. In total, 12.1%

(11.9–12.2%) of all tests done in the unmeasured urine albumin cohort are misclassified, 13.1% (13.0–13.3%) in patients with normoalbuminuria, 14.5% (14.2–14.8%) in patients with microalbuminuria, and 14.7% (14.1–15.3%) in patients with macroalbuminuria.

Mean sojourn time, i.e. the average time spent in each state, decreased with increasing severity of eGFR and albuminuria stage (Table 3). One exception was for macroalbuminuric patients in eGFR stage G5, for whom the mean sojourn time was greater than for microalbuminuric patients in eGFR stage G5. However, few patients were present in the more severe diseases states and the 95% confidence intervals of the two estimates substantially overlap.

Discussion

We have developed a statistical model for kidney function monitoring over time, using a large clinical database of longitudinal kidney function measurements from an unselected primary care cohort. This model takes into account that observed kidney

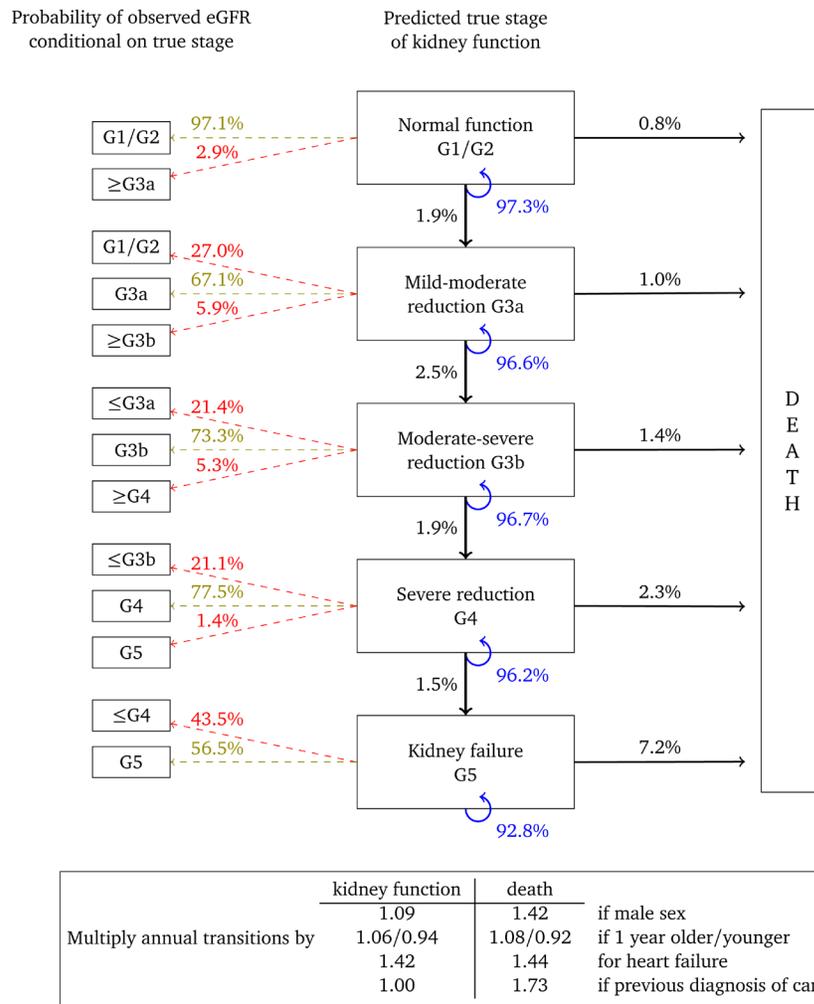


Figure 3. Annual transition model diagram for patients with normoalbuminuria at baseline. Probabilities are based on a woman aged 60, without heart failure or a previous diagnosis of cancer.

function is measured with error and uses statistical methodology to estimate the underlying ‘true’ rate of progression. We stratified our models by albuminuria stage in accordance with the findings of previous studies that showed that urine albumin excretion is a significant risk factor for the progression of CKD and the development of ESRD¹⁹⁻²¹. Our analyses suggest that kidney function declines more rapidly in men than in women, independent of other risk factors. Existing evidence for differences in the rates of progression between men and women is conflicting^{3,22,23}. Our analysis supports the observations of others, that men are over-represented in the latter stages of CKD²⁴, with our model predicting a slower progression of kidney disease for women in the unmeasured urine albumin and normoalbuminuria cohorts. The fact that women are over-represented at CKD stage 3 may be due to the fact that women tend to live longer than men.

We estimated the probability of misclassification conditioning on true eGFR stage. A consistent pattern is seen across the

different baseline urine albumin levels and by eGFR stage. Our model suggests that on average, change in underlying kidney function is slow with mean sojourn times in stage G3a and G3b being between 15 and 25 years for patients without elevated urine albumin. Given the slow rate of change and the high chance that observed eGFR misclassifies the true eGFR stage, frequent testing of eGFR in these populations will inevitably lead to the detection of more spurious change than real change.

We assessed whether our models of kidney disease progression would be improved by adjusting for clinical characteristics that were *a priori* considered to be associated with increased risk, and therefore, faster progression. Our analysis did not support the notion that diabetes, hypertension, peripheral vascular disease, ischaemic heart disease, stroke or transient ischaemic attack are independently associated with deterioration of kidney function once albuminuria stage and updated eGFR are accounted for. We conclude that conditioning on eGFR stage

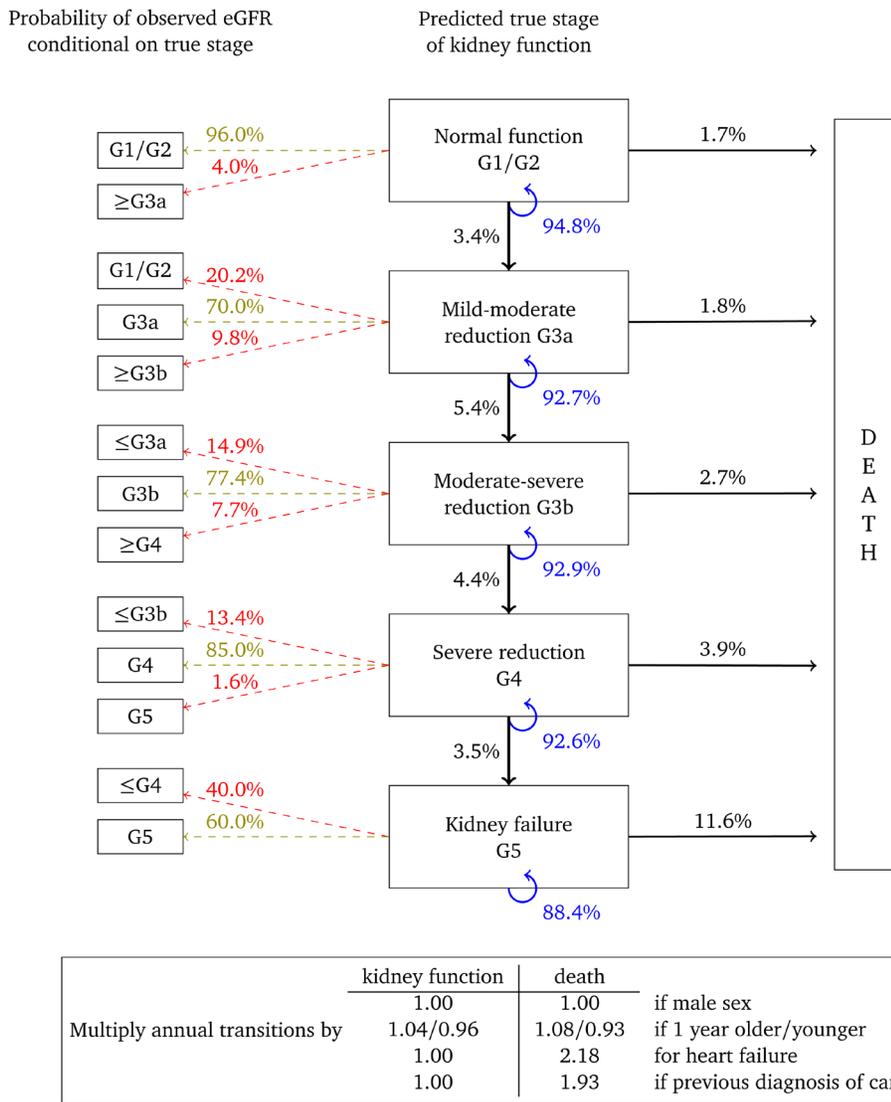


Figure 4. Annual transition model diagram for patients with microalbuminuria at baseline. Probabilities are based on a woman aged 60, without heart failure or a previous diagnosis of cancer.

and urine albumin levels, knowledge of diabetes status is less important, but we cannot rule out that our study may be under-powered to detect small but real effects on transition rates.

A major strength of this study is that we have taken a very large and unselected sample of patients from a database that has been shown to be representative of the wider UK population⁷. Our model for progression takes into account multiple stages of kidney function and the competing risk of death from any cause. We have also employed a method that takes into account that eGFR is observed with error, and simultaneously estimates true underlying eGFR²⁵. This means that we can estimate misclassification probabilities and evaluate the effects of different monitoring strategies. We used a split-sample

approach to assess for potential over-fitting and the internal validity of the model.

Our study has a number of limitations. Our data was not collected for the purpose of conducting a study about modelling progression of kidney function. As a consequence, we do not know the reasons tests were conducted, and for many patients, records were incomplete and examination times were irregular. The extent to which this could bias our findings is unclear as it depends on our understanding of the examination scheme used by the doctors. We recognise three potential mechanisms for these tests to occur in a primary care setting. A significant number of creatinine tests will be ‘random’ with respect to the kidney function, because they would have been ordered

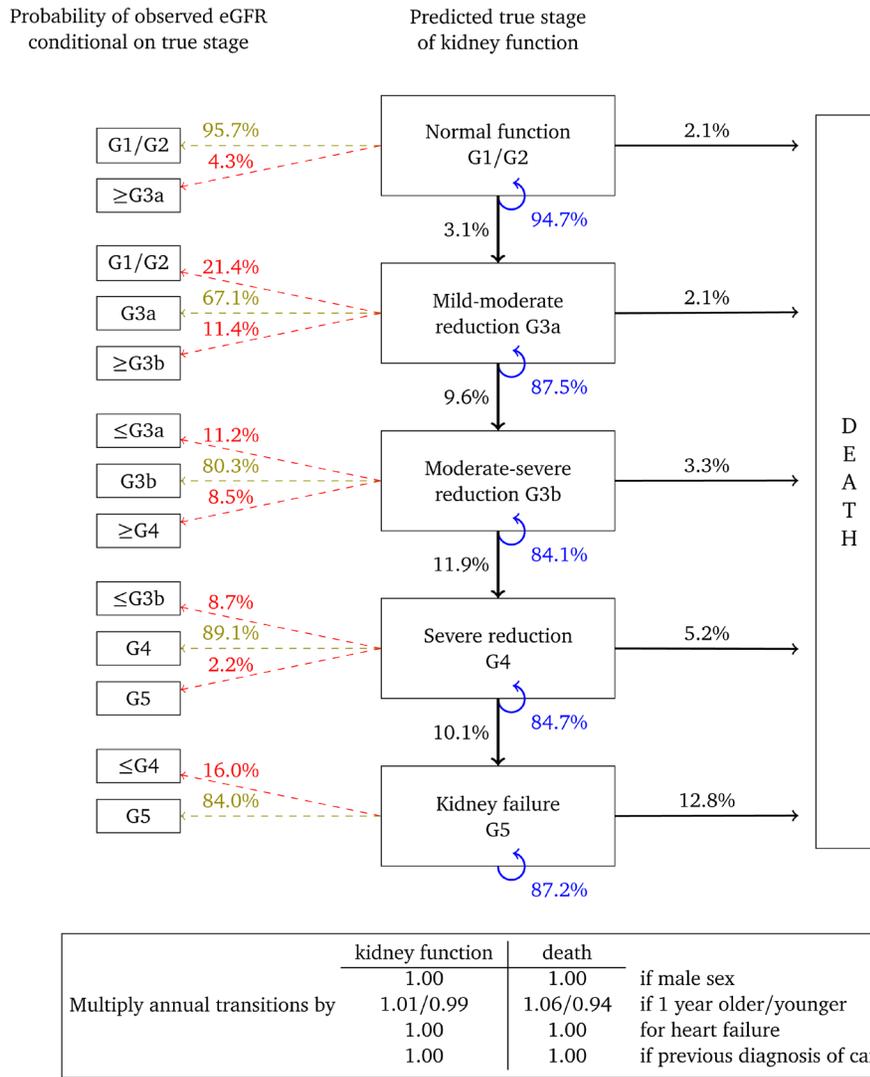


Figure 5. Annual transition model diagram for patients with macroalbuminuria at baseline. Probabilities are based on a woman aged 60, without heart failure or a previous diagnosis of cancer.

Table 2. Probability (%) that any eGFR test is under-graded and/or over-graded, by albuminuria stage. 95% confidence intervals shown in brackets.

Grading	Unmeasured	Normoalbuminuria	Microalbuminuria	Macroalbuminuria
Under-grading	8.3 (8.2–8.5)	9.2 (9.0–9.3)	8.7 (8.5–8.9)	8.5 (8.1–9.0)
Over-grading	3.7 (3.6–3.8)	4.0 (3.9–4.0)	5.8 (5.6–6.0)	6.2 (5.8–6.6)
Total	12.1 (11.9–12.2)	13.1 (13.0–13.3)	14.5 (14.2–14.8)	14.7 (14.1–15.3)

as part of a routine check-up and not specifically to monitor or diagnose kidney disease. This could be a result of the co-reporting of serum creatinine as part of ‘test batches’ in which other biomarkers would have been of primary interest, or because serum creatinine may have been requested

prior to the initiation of a potentially nephrotoxic drug. For some patients, the timing of the next measurement will have been influenced by the current kidney function level. This is likely to have happened if the purpose of the test is to monitor CKD and current clinical guidelines are followed¹. This

Table 3. Mean sojourn times, by albuminuria and eGFR stage. 95% confidence intervals shown in brackets.

eGFR Stage	Unmeasured	Normoalbuminuria	Microalbuminuria	Macroalbuminuria
G1/2	30.5 (29.2–31.8)	20.1 (19.3–20.9)	10.4 (9.7–11.1)	13.1 (11.3–15.3)
G3a	26.7 (25.2–28.4)	15.7 (14.9–16.6)	7.5 (6.9–8.1)	6.0 (5.1–7.0)
G3b	25.1 (23.3–27.1)	15.7 (14.6–16.8)	7.3 (6.7–8.1)	4.5 (3.9–5.3)
G4	19.5 (17.4–21.8)	12.8 (11.5–14.3)	6.8 (5.9–7.7)	4.5 (3.7–5.5)
G5	7.0 (5.6–8.9)	5.9 (4.6–7.6)	3.6 (2.8–4.7)	4.2 (3.1–5.7)

mechanism has been referred to as ‘doctor’s care’ in the literature. The third scenario is when a patient initiates the timing of their test themselves, so called ‘patient self-selection’. Of the three scenarios, we consider the self-selection scenario possible but less likely than the other schemes due to the asymptomatic nature of kidney function loss in all but the end-stages of the disease. Gröger *et al.*²⁶ showed that estimated transition rates are only biased under the “patient self-selection” examination schemes and transition rate estimates are unbiased if inefficient under doctor’s care scheme. In the case of random timing, the estimates are both efficient and unbiased.

We attempted to include a state to represent transient and acute loss of kidney function (acute kidney injury) as this is a contributing factor to CKD, but the addition of this non-absorbing state with pathways back to each state resulted in over-parametrisation of the model. Furthermore, data on urine albumin, body mass index and ethnicity is missing in a large number of patients in CPRD. To overcome this, we created a sub-group of patients in whom urine albumin was not recorded. The omission of ethnicity in this model is a limitation as kidney function decline is considered to differ between ethnic groups. We were not able to adjust our models for ethnicity, as historically, ethnicity has been poorly recorded in CPRD.

It is likely that once a patient’s kidney function has been observed in stage 4 or 5, they are referred to specialist care, with subsequent kidney function testing occurring outside the CPRD database. Hence, these patients’ records are missing from our study, which potentially explains why transition rates slow down rather than increase, as might be expected. In a study of electronic health records data from Pennsylvania, a similar model was fit to eGFR records, and reported that transition probabilities between kidney function stages generally increased as stage increased for all but stage 3²⁵. Even so, our model calibrates well with reports of progression to ESRD from different stages. For example, Tangri *et al.*²⁷ reported that three from 2,014 people with CKD stage 3 at baseline progressed to ESRD after three years of follow-up. Assuming this population contained an equal proportion of people with CKD stage 3a and 3b, then our model, based on the unmeasured urine albumin cohort, would predict that just one person would reach stage 5 after three years. Using the model for patients with normoalbuminuria, it would be three people. From the

same study, 22 of 826 people progressed from stage 4 at baseline to kidney failure after three years. Our models predict 25 people with unmeasured urine albumin and 46 people with normoalbuminuria would reach stage 4. In a study reporting on sex differences in CKD progression, the rate of ESRD per 100 person-years was 3.1 in women and 3.8 in men. Based on our model for patients with normoalbuminuria our equivalent estimates are 1.9 and 2.3, but 2.07 and 2.13 for patients with microalbuminuria and 3.0 and 3.2 for patients with macroalbuminuria. Our study shows that kidney function deteriorates slowly in most patients with average sojourn times in decades rather than years. Whilst eGFR is widely used to measure kidney function we estimate that the potential for misclassification is large and clinically relevant, with implications for monitoring for rapid kidney function loss or pharmacovigilance. For example, of 1,741 people with CKD stage 3 recruited for a study from 32 primary care practices in the UK²⁸, 496 were in remission at baseline (although qualifying at the recruitment stage) and of these, 157 were back to CKD stage 3 at one year, with a further 132 returning to stage 3 CKD by five years. This type of pattern is consistent with our model, in which underlying kidney function only deteriorates but is observed with error. If our model is correct, then it is clear to see how monitoring CKD periodically will confuse and might lead to inappropriate action.

Conclusions

We have developed a model to predict decline in kidney function and used it to assess different monitoring strategies and screening programmes. The model takes into account stage progression and test error, which were recently identified as important for future economic evaluations of CKD testing²⁹. Future work in this field could look to validate this model in another primary care population, ideally one in which patients are followed throughout including stages 4 and 5.

Data availability

Underlying data

The data used in this study are not publicly available and were obtained under licence. The terms of this license do not permit us to share the data. However, those wishing to replicate our analysis in this database can apply directly to the Medicines and Healthcare Products Regulatory Agency (MHRA) for access to the CPRD, at enquiries@cprd.com. The conditions

under which the MHRA will grant access are beyond our control, but are explained at <https://www.cprd.com/research-applications>.

Extended data

Figshare: Statistical models for the deterioration of kidney function in a primary care population: A retrospective database analysis (Extended Data). <https://doi.org/10.6084/m9.figshare.9741611.v1>¹⁴.

This project contains the following extended data:

- Extended Data.pdf (document containing Tables S1–20 and Figures S1–5)

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/) (CC-BY 4.0).

Software availability

Source code available from: <https://github.com/OxPrimaryCareStats/egfr-decline/tree/v1.0.0>

Archived source code at time of publication: <https://doi.org/10.5281/zenodo.3377113>¹⁸

License: MIT License

Acknowledgements

We would like to thank Alice Fuller and Dr Sarah Lay-Flurrie for their hard work in providing much of the initial data management for this project.

References

- National Institute for Health and Care Excellence: **Chronic kidney disease in adults: assessment and management: Clinical guideline [CG182]**. Technical report, National Institute for Health and Care Excellence, 10 Spring Gardens, London, SW1A 2BU. 2014.
[Reference Source](#)
- Weiner DE, Tighiouart H, Amin MG, *et al.*: **Chronic kidney disease as a risk factor for cardiovascular disease and all-cause mortality: a pooled analysis of community-based studies.** *J Am Soc Nephrol.* 2004; 15(5): 1307–15.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Eriksen BO, Ingebreten OC: **The progression of chronic kidney disease: a 10-year population-based study of the effects of gender and age.** *Kidney International.* 2006; 69(2): 375–382.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Lamb E, Brettell EA, Cockwell P, *et al.*: **The eGFR-C study: accuracy of glomerular filtration rate (GFR) estimation using creatinine and cystatin C and albuminuria for monitoring disease progression in patients with stage 3 chronic kidney disease—prospective longitudinal study in a multiethnic population.** *BMC Nephrol.* 2014; 15: 13.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Spence D: **Bad medicine: chronic kidney disease.** *BMJ.* 2010; 340: c3188.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Fink H, Ishani A, Taylor B, *et al.*: **Screening for, monitoring, and treatment of chronic kidney disease stages 1 to 3: a systematic review for the U.S. Preventive Services Task Force and for an American College of Physicians Clinical Practice Guideline.** *Ann Intern Med.* 2012; 156(8): 570–581.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Herrett E, Gallagher AM, Bhaskaran K, *et al.*: **Data resource profile: Clinical Practice Research Datalink (CPRD).** *Int J Epidemiol.* 2015; 44(3): 827–836.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- National Kidney Foundation: **K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification.** *Am J Kidney Dis.* 2002; 39(2 Suppl 1): S1–266.
[PubMed Abstract](#)
- Baum L, Petrie T: **Statistical Inference for Probabilistic Functions of Finite State Markov Chains.** *Ann Math Stat.* 1966; 37(6): 1554–1563.
[Publisher Full Text](#)
- Baum L, Eagon J: **An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology.** *Bull Am Math Soc.* 1967; 73(3): 360–363.
[Publisher Full Text](#)
- Baum L, Sell G: **Growth transformations for functions on manifolds.** *Pac J Math.* 1968; 27(2): 211–227.
[Publisher Full Text](#)
- Baum L, Petrie T, Soules G, *et al.*: **A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains.** *Ann Math Stat.* 1970; 41(1): 164–171.
[Publisher Full Text](#)
- Baum L: **An inequality and associated maximization technique in statistical estimation of probabilistic functions of a markov process.** *Inequalities.* 1972; 3: 1–8.
[Reference Source](#)
- Oke JL, Feakins B, Schlackow I, *et al.*: **Statistical models for the deterioration of kidney function in a primary care population: A retrospective database analysis (Extended Data).** 2019; 8.
<http://www.doi.org/10.6084/m9.figshare.9741611.v1>
- Viterbi A: **Error bounds for convolutional codes and an asymptotically optimum decoding algorithm.** *IEEE Trans Inf Theory.* 1967; 13(2): 260–269.
[Publisher Full Text](#)
- R Core Team: **R: A Language and Environment for Statistical Computing.** R Foundation for Statistical Computing, Vienna, Austria, 2019.
[Reference Source](#)
- Jackson CH: **Multi-state models for panel data: The msm package for R.** *J Stat Softw.* 2011; 38(8): 1–29.
[Publisher Full Text](#)
- Feakins B: **Oxprimarycarestats/egfr-decline: Initial release.** 2019.
<http://www.doi.org/10.5281/zenodo.3377113>
- Chang WX, Arai S, Tamura Y, *et al.*: **Time-dependent risk factors associated with the decline of estimated GFR in CKD patients.** *Clin Exp Nephrol.* 2016; 20(1): 58–70.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Inaguma D, Imai E, Takeuchi A, *et al.*: **Risk factors for CKD progression in Japanese patients: findings from the Chronic Kidney Disease Japan Cohort (CKD-JAC) study.** *Clin Exp Nephrol.* 2017; 21(3): 446–456.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Coresh J, Heerspink HJL, Sang Y, *et al.*: **Change in albuminuria and subsequent risk of end-stage kidney disease: an individual participant-level consortium meta-analysis of observational studies.** *Lancet Diabetes Endocrinol.* 2019; 7(2): 115–127.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ricardo AC, Yang W, Sha D, *et al.*: **Sex-Related Disparities in CKD Progression.** *J Am Soc Nephrol.* 2019; 30(1): 137–146.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Neugarten J, Acharya A, Silbiger SR: **Effect of gender on the progression of nondiabetic renal disease: a meta-analysis.** *J Am Soc Nephrol.* 2000; 11(2): 319–329.
[PubMed Abstract](#)
- Glasscock R, Winearls C: **Screening for CKD with eGFR: doubts and dangers.** *Clin J Am Soc Nephrol.* 2008; 3(5): 1563–1568.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Luo L, Small D, Stewart WF, *et al.*: **Methods for estimating kidney disease stage transition probabilities using electronic medical records.** *EGEMS (Wash DC).* 2013; 1(3): 1040.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Grüger J, Kay R, Schumacher M: **The validity of inferences based on incomplete observations in disease state models.** *Biometrics.* 1991; 47(2): 595–605.
[PubMed Abstract](#) | [Publisher Full Text](#)

27. Tangri N, Inker LA, Hiebert B, *et al.*: **A Dynamic Predictive Model for Progression of CKD.** *Am J Kidney Dis.* 2017; **69**(4): 514–520.
[PubMed Abstract](#) | [Publisher Full Text](#)
28. Shardlow A, McIntyre NJ, Fluck RJ, *et al.*: **Chronic Kidney Disease in Primary Care: Outcomes after Five Years in a Prospective Cohort Study.** *PLoS Med.* 2016; **13**(9): e1002128.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
29. Sutton AJ, Breheny K, Deeks J, *et al.*: **Methods Used in Economic Evaluations of Chronic Kidney Disease Testing - A Systematic Review.** *PLoS One.* 2015; **10**(10): e0140063.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status: ?

Version 1

Reviewer Report 11 September 2019

<https://doi.org/10.5256/f1000research.22225.r53695>

© 2019 Nitsch D. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Dorothea Nitsch

Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine (LSHTM), London, UK

The authors used CPRD records to inform a multistate model on progression of kidney disease. They appropriately considered that there could be measurement error and allowed for this to some extent.

They used outcome variables that have not been validated for incident analyses in CPRD(dialysis/transplantation).

The underlying model for 'true' but unobserved kidney disease assumed that kidney disease could only get worse over time but not better - but this is not supported by actual data which suggest that transient decreases can improve over time and may not necessarily need to be permanent (though such changes may be a risk marker for later decline). Here the authors should discuss more about the chronicity assumption within the CKD definition and how they parametrised this in their data.

Then there was a model for measurement error, but this seems to be a static model, i.e. assuming that measurement error does not vary over time - did I understand this correctly? Because over time in the UK the way creatinine gets measured and reported has changed dramatically during the study period. This change in reporting of calibrated creatinines in effect hides progression over time when crude analyses are used as different labs shifted to calibration and reporting of creatinine to IDMS at different points in time - thereby shifting the entire creatinine distribution down by 5% over the years prior to 2014 (i.e. hiding a decrease in kidney function over time).

Did the authors recalculate eGFR from the creatinines (which then requires some thought about time-dependent measurement error which varies by lab and time) - or use reported eGFRs (which then means they lost a lot of measurements based on thresholds with informative missingness as eGFR does not get reported uniformly - some labs only report values in a range of >15 and <60, others report up to 90ml/min/1.73m2)? Dependent on whether the authors used creatinine or eGFR they have discussed further biases in their design and explicitly allow for such biases in their model.

Then there is the overall study design which is a somewhat odd cohort as it is dependent on having three

or more eGFR (or creatinine?) tests and that is not representative of the general population - the survey and CPRD validated prevalence of reduced eGFR in the UK population is about 6% and here the numbers are much higher (3-7 fold depending on albuminuria category) simply because these represent an enriched sample as GPs had a reason to test more than once but indeed three times. Does this enriched sample really represent 'a model for kidney disease progression'? The authors should discuss this and whether people who should have been tested but weren't tested may be a high risk group. The sample here represents a group of patients who engage with the health service but the people who progress in truth may be not fully captured. This needs to be discussed more.

There were real financial reasons for testing e.g. annual testing for diabetes (started in 2004), and less so for other illnesses, but risk factors for CKD determine testing rates and especially repeat testing as reported by the UK National CKD Audit, and there is some understanding from this audit how testing schemes are carried out. So I would disagree with "A major strength of this study is that we have taken a very large and unselected sample of patients from a database that has been shown to be representative of the wider UK population" as stated in the discussion - this is a selected sample and not representative of what happens overall in terms of kidney function decline as not all are tested the same way. I would have stratified by underlying comorbidity and not simply adjusted for it.

I totally agree with the authors about the selective loss to follow-up with loss of people who are managed by other specialities including renal in secondary care.

Overall this is an interesting analysis, but more work is needed to convince me that this model should be used for economic modelling.

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

No

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

No

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Renal epidemiology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research