Check for updates

RESEARCH ARTICLE

REVISED Statistical models for the deterioration of kidney function in a primary care population: A retrospective database analysis [version 2; peer review: 2 approved]

Jason L Oke[1], Benjamin G Feakins[1], Iryna Schlackow[2], Borislava Mihaylova[2,3], Claire Simons[2], Chris A O'Callaghan[4], Daniel S Lasserson[5], F D Richard Hobbs[1], Richard J Stevens[1], Rafael Perera[1]

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

v2 **First published:** 10 Sep 2019, **8**:1618
https://doi.org/10.12688/f1000research.20229.1
Latest published: 08 Aug 2022, **8**:1618
https://doi.org/10.12688/f1000research.20229.2

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

Open Peer Review

Approval Status ✔ ✔

	1	2
version 2 (revision) 08 Aug 2022	✔ view	✔ view
version 1 10 Sep 2019	? view	? view

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

2. **Lawrence Hunsicker**, University of Iowa, Iowa City, USA

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

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 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 (cookiecutter@posteo.eu)

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: © 2022 Oke JL *et al*. This is an open access article 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.

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 2; peer review: 2 approved]** F1000Research 2022, **8**:1618 https://doi.org/10.12688/f1000research.20229.2

First published: 10 Sep 2019, **8**:1618 https://doi.org/10.12688/f1000research.20229.1

REVISED **Amendments from Version 1**

In response to the Reviewers' comments, we have added a paragraph to the Discussion explaining the potential impact of outcome measure categorisation and Markov model assumptions. Regarding the loss of information, it is typically small when the number of categories is large (as was the case in our study). As for potential violations of the Markov assumption, these were mitigated through the stratification of our models by baseline albuminuria stage (a key factor determining trajectory of kidney function decline) and through conditioning the models on updated covariates.

One of the key assumptions of our models, which was not previously highlighted, was that kidney function could only deteriorate with age. We have added a sentence to our Discussion to highlight this.

We have also edited the Discussion to clarify our study's eligibility criteria and to explain that the results are generalisable to the UK general practice population (as opposed to the UK population). Serum creatinine tests are commonly performed in UK general practice, and not necessarily for the purpose of monitoring kidney function. Therefore, our study was able to capture an unselected UK general practice population, but may have omitted patients with end-stage renal disease, or patients who did not engage with their general practice.

Any further responses from the reviewers can be found at the end of the article

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 reninangiotensin-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 population[2]. Yet, kidney function declines slowly with age and ESRD is rare, even for people with moderately impaired renal function (eGFR 30–59 ml/min/1.73m^2). 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 people[3]. 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 benefit[5].

There have been no trials of screening or monitoring for CKD[6] 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 Adisory 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 2002[8] 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/1.73m^2 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, 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 algorithm[15] 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* package[17]. 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

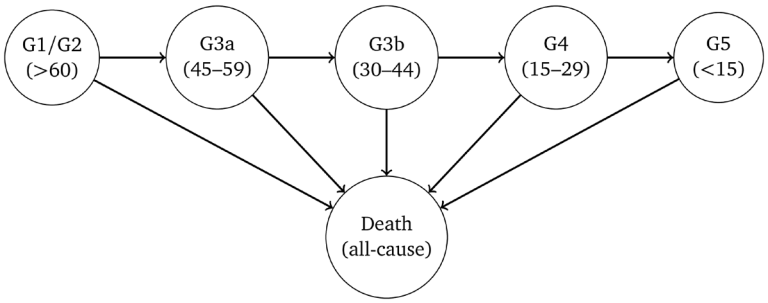


Figure 1. Representation of the model for the deterioration of kidney function over time. Arrows indicate permitted (instantaneous) transitions. The numbers in brackets depict the estimated glomerular filtration rate ranges (in ml/min/1.73^2) associated with each stage.

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%)

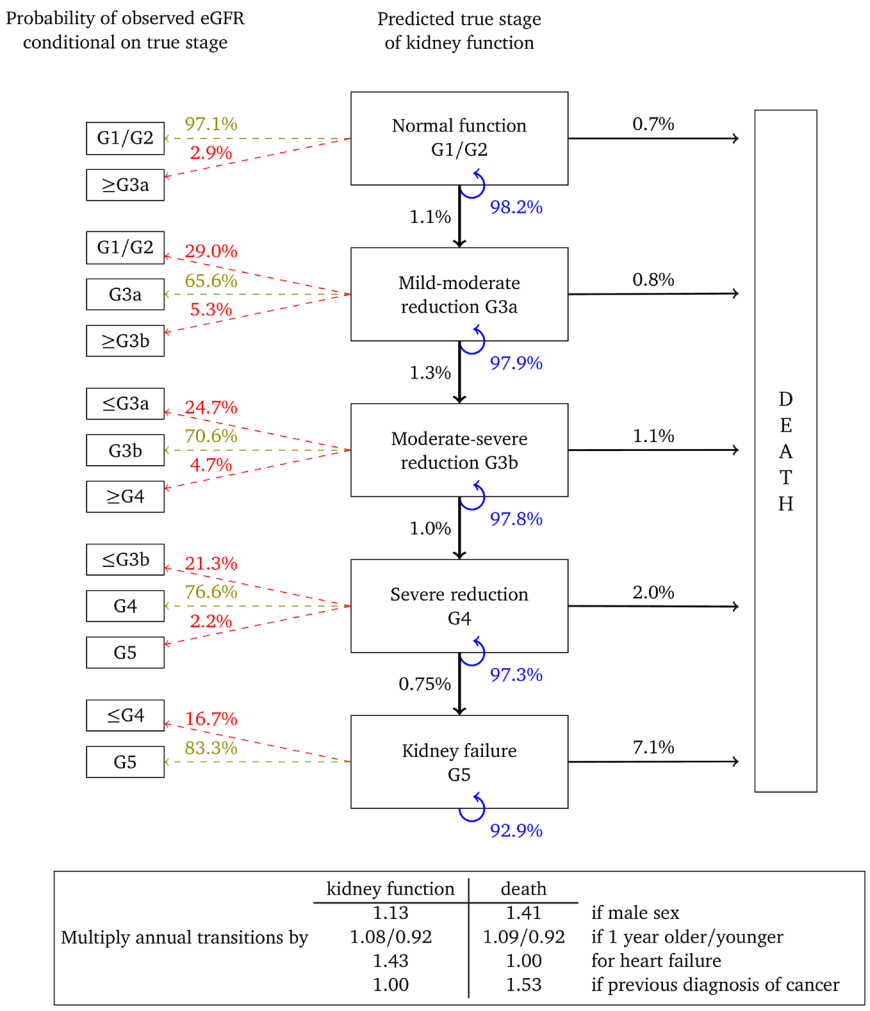


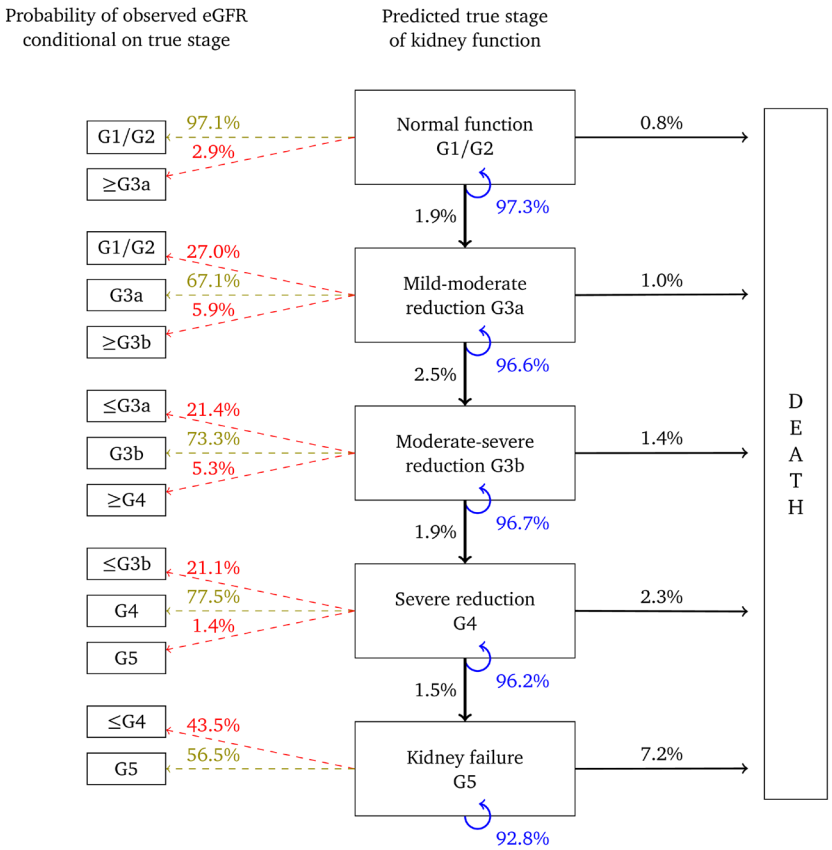
Figure 2. Annual transition model diagram for patients with unmeasured urine albumin at baseline. Probabilities are based on a woman aged 60, without heart failure or a previous diagnosis of cancer.

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 $1.0 \times 1.08^{20} = 4.66\%$ 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 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



	kidney function	death	
Multiply annual transitions by	1.09	1.42	if male sex
	1.06/0.94	1.08/0.92	if 1 year older/younger
	1.42	1.44	for heart failure
	1.00	1.73	if previous diagnosis of cancer

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.

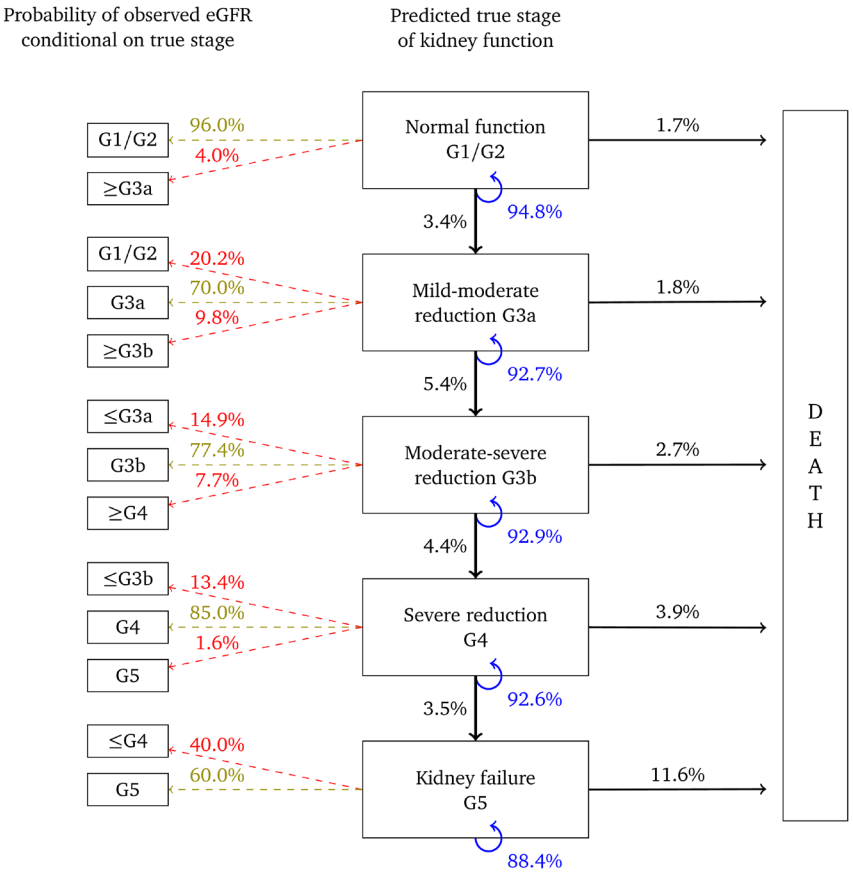
with normoalbuminuria, 14.5% (14.2–14.8%) in patients with microalbuminuria, and 14.7% (14.1–15.3%) in patients with macroalbuminuria.

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 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 ESRD[19–21]. 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 CKD[24], 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.



	kidney function	death	
Multiply annual transitions by	1.00	1.00	if male sex
	1.04/0.96	1.08/0.93	if 1 year older/younger
	1.00	2.18	for heart failure
	1.00	1.93	if previous diagnosis of cancer

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.

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 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 primary care population[7]. Inclusion into the study was conditional upon having three or more serum creatinine measurements, but creatinine is commonly measured in UK general and not necessarily for the purpose of monitoring kidney function or diagnosing kidney disease. 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

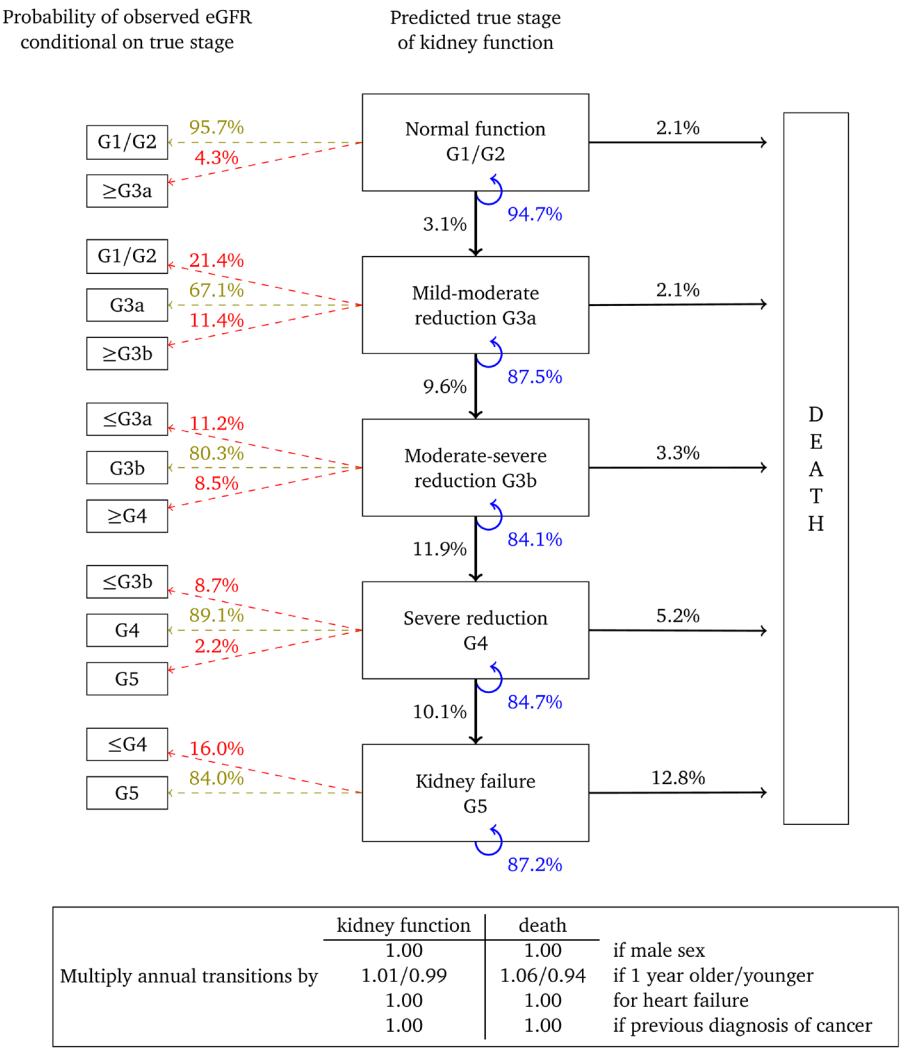


Figure 5. Annual transition model diagram for patients with macroalbuminuria at baseline. Probabilities are based on a woman aged 60, without heart failure or a previous diagnosis of cancer.

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)

takes into account that eGFR is observed with error, and simultaneously estimates true underlying eGFR[25]. 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

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)

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 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 followed[1]. This 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. Grüger *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 could be criticised for using an approach that categorises kidney function rather than a method that models continuous eGFR, such as generalised linear mixed models. This is because categorisation can lead to loss of information and reduced statistical power. However, such information loss is typically small when the number of categories is large[27], as was the case in our study. Furthermore, the use of categories that naturally aligned with clinically meaningful eGFR stages added to the interpretability of our findings. In addition, HMMs assume that all individuals within a state are interchangeable, and that the chance of progression to subsequent states depends only on the current state. This assumption may not hold if a patient whose kidney function has previously rapidly declined continues along this trajectory. To mitigate this, we have included updated risk factor information in our model and stratified by baseline albuminuria status. Further research could include assessing the impact of the Markov assumption on predicting kidney function decline using HMMs.

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. Our study design means that we would also miss patients with ESRD who had not engaged in primary care. 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 3[25]. Even so, our model calibrates well with reports of progression to ESRD from different stages. For example, Tangri *et al.*[28] 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 UK[29], 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. The assumption that true underlying kidney function only deteriorates with age is a fundamental part of the model and further research could investigate alternative models for underlying kidney progression and their impact monitoring recommendations.

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 testing[30]. 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 (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

1. 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
2. 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
3. Eriksen BO, Ingebretsen OC: **The progression of chronic kidney disease: a 10-year population-based study of the effects of gender and age.** *Kidney Int.*. 2006; **69**(2): 375–382.
PubMed Abstract | Publisher Full Text
4. 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
5. Spence D: **Bad medicine: chronic kidney disease.** *BMJ.* 2010; **340**: c3188.
PubMed Abstract | Publisher Full Text
6. 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
7. 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
8. 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
9. 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
10. 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
11. Baum L, Sell G: **Growth transformations for functions on manifolds.** *Pac J Math.* 1968; **27**(2): 211–227.
Publisher Full Text
12. 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
13. 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

14. 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

15. 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

16. R Core Team: **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing, Vienna, Austria, 2019.
Reference Source

17. Jackson CH: **Multi-state models for panel data: The msm package for R.** *J Stat Softw.* 2011; **38**(8): 1–29.
Publisher Full Text

18. Feakster: **Oxprimarycarestats/egfr-decline: Initial release**. 2019.
http://www.doi.org/10.5281/zenodo.3377113

19. 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

20. 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

21. 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

22. 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

23. 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

24. Glassock 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

25. 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

26. 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. Altman DG, Royston P: **The cost of dichotomising continuous variables.** *BMJ.* 2006; **332**(7549): 1080.
PubMed Abstract | Publisher Full Text | Free Full Text

28. 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

29. 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

30. 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 2

Reviewer Report 04 October 2022

https://doi.org/10.5256/f1000research.136788.r146903

© **2022 Hunsicker L.** 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.

✔ **Lawrence Hunsicker**
Department of Internal Medicine, University of Iowa, Iowa City, USA

The revised version includes no new or revised analyses. So this Version 2 report deals entirely with the authors' responses to my Version 1 Report and the edits in their Discussion. I repeat the positive aspects of this manuscript summarised in the first two paragraphs of my Version 1 Report, dealing with the importance of the issue raised by the authors, the almost total absence of evidence on the utility of measurement of eGFR and proteinuria in a population of subjects similar to those followed in UK General Practices, the availability of the nearly unique set of data from the Clinical Practice Research Datalink, and the authors' competence in performing their Hidden Markov Model analysis. My reservations dealt mostly with the authors' choice of analytic model.

The authors' rationale for their use of their Hidden Markov Model is that it permits adjustment for the well-documented degree of CKD stage misclassification that is revealed when estimates of eGFR are compared with directly measured GFR (itself measured with significant "error" when evaluated by replicate measurements). But their Hidden Markov Model also makes assumptions that may well not hold in reality. There are two specific issues with the authors' Markov assumptions: a. that eGFR once "lost" cannot be recovered, and b: that the history of prior changes in renal function is not informative once a Markov transition has occurred. This model seems to be based on an assumption that reduction of eGFR is mostly associated with irreversible changes in renal structure or physiology. This assumption is particularly risky in subjects with relatively well preserved renal function and absent or trace proteinuria -- demonstrably the large majority of subjects in the authors' cohort and evidently the primary focus of their broader project. In these subjects, the likelihood of a transient reduction of eGFR due to any episode leading to dehydration may well be greater than of an event causing an irreversible loss of function (conflicting with the assumption that "true" [hidden] eGFR cannot increase following a drop in eGFR). Further, subjects with prior episodes of AKI are known to have greater subsequent chance of progressing to ESRD, and subjects receiving treatment with any of the three classes of treatment shown to be effective in delaying progression (renin/angiotensin/aldosterone inhibitors, reduction of systolic blood pressure, and more recently SGLT2 inhibitors) are all associated with immediate reductions of eGFR that are associated with later slower risk of subsequent progression. These facts conflict with the assumption that past changes in eGFR

leading to a Markov stage transition are not informative of the chance of subsequent loss of eGFR. I suggested using an alternative model -- joint modeling -- that deals similarly with the issue of errors in measurement of proposed predictors by estimating a true but "hidden" path of eGFR over time from the observed values and uses estimates from that model at relevant time points as the estimated "true" current value for the joint survival model. This technique permits use of all the continuous primary data without collapsing it into categories, makes no assumption excluding increases of eGFR, and permits many options for assumptions about the impact of earlier values. Additionally, this approach would permit direct evaluation of the concordance and the rate of classification error of the relevant primary outcomes -- ESRD or death. In my Version 1 report, I noted that "Given the huge amount of information available for this analysis, the inefficiency of the hidden Markov model may be overcome in the analysis of the overall community estimates of rate of progression." This seems to be the case in the paper cited in their response to my critique, and I agree that their analytic method is useful for evaluation of choice of overall or group policy. I suspect, though, that alternative analytic methods permitting use of all of the continuous data might be more powerful in refining predictions, and therefore permitting individualized recommendations or guidelines applicable to specific patients.

In their response to my Version 1 report, the authors' have included new text recognizing the limitations of the Markov approach and recognized that "Further research could include assessing the impact of the Markov assumption on predicting kidney function decline using HMMs." Given this addition, I now approve this manuscript.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Nephrology, Epidemiology, Statistics, Clinical trial design, performance, and analysis.

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.

Reviewer Report 23 August 2022

https://doi.org/10.5256/f1000research.136788.r146904

© 2022 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

I have no further comments.

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.

Version 1

Reviewer Report 28 August 2020

https://doi.org/10.5256/f1000research.22225.r69718

© 2020 Hunsicker L. 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.

?

Lawrence Hunsicker
Department of Internal Medicine, University of Iowa, Iowa City, USA

The authors of this manuscript report on the long-term rate of progression of chronic kidney disease (CKD) in a very large cohort of UK general practice patients followed as part of the Clinical Practice Research Datalink in the time between 1 April 2005 and 31 March 2014. Progression of kidney disease over each patient's course was assessed as the rate of progression from one stage to the next across the six currently standard KDOQI/KDIGO "stages" of renal insufficiency (Stages 1, 2, 3a, 3b, 4, and 5), which are defined by ranges of estimated glomerular filtration rate (eGFR). The rate of progression from one stage to the next was modeled as a Hidden Markov Model, which assumes that subjects are at "true" eGFR stages at each evaluation point, but that these stages may be misclassified based on random error of the eGFR estimate, so that the "observed" transitions may be misidentified. The overall conclusions of the authors are: a) that progression of renal disease to renal failure occurs quite slowly in general, with only a small fraction of patients that progress to kidney failure, and b) that progression of kidney disease as evidenced by [exclusive] use of progression in the KDOQI/KDIGO Stage of CKD, though generally reflective of the average course of kidney disease in the community, is associated with significant error when applied to an individual. Though the authors don't say this, the implication is that use of their analysis may not be very helpful in establishing guidelines as to in whom, and at what repeated intervals, eGFR determinations would be cost effective.

The major strength of this study is the almost unique test bed offered by access to the data of almost 2,000,000 UK qualified general practice patients from Clinical Practice Research Datalink with "up-to-standard" data collection. Given the choice of analytic method, the authors' analysis is technically entirely competent. The assumptions and methods are well described. The results are clearly explained and put in appropriate context in the Discussion Section. The major limitation of the available data is that patients with the more advanced Stages of kidney disease are likely to have been referred to specialist providers, so that data on the more advanced stages of kidney disease, where progression is generally both more rapid and more clinically important, are more limited and more susceptible to bias. The authors clearly acknowledge this. Further, the issue addressed by this manuscript is potentially important, at least from a health economics point of

view. Current policy recommendations uniformly encourage periodic assessment of renal function "in people with, or at risk of, chronic kidney disease (CKD)." But as noted above, in general kidney disease progresses quite slowly and only a small minority of subjects even with documented kidney disease ever progress to chronic renal failure. There is almost no empirical data upon which to make any data-based recommendation about in whom or how often the "periodic assessments" should be done.

But the choice of analytic method is subject to criticism. The continuous variable eGFR may range from over 100 (mL/min/1.73 M^2) to close to 0. The reduction of this range of values to a classification with 6 levels is associated with massive information loss. Markov models deal with chances of transition from one state to another, and all members of one state are considered to be interchangeable, so that a patient with an eGFR of 100 (at Stage 1/2 with the combination of these two stages) is treated as equivalent to one with an eGFR of 62. But surely the chance of progression from an eGFR of 100 to an eGFR < 60 (Stage 3a) is much lower than the chances of progression from a value of 62 to < 60 . Similarly, it is intuitively likely (though not at all certain) that a subject with a history of rapid loss of eGFR in the past will have more rapid progression in the future. But with a Markov model, all history is lost, and only the current stage, and not the prior history is taken into account. Given the huge amount of information available for this analysis, the inefficiency of the hidden Markov model may be overcome in the analysis of the overall community estimates of rate of progression. But these results will be of almost no use in prediction of the progression of kidney disease in the individual patient. And it is the individual patient's progress, not the average rate of progression in the community, that we are trying to clarify in recommending individual testing eGFR and setting repeat testing intervals.

An alternative approach would have been to use a hierarchical ("mixed") analytic method (or perhaps even better, joint analysis combining the hierarchical analysis with a time to event analysis for ESRD and death) to evaluate the overall community rate of progression while estimating the distribution of individual rates of progression. Mixed models deal automatically with the problem that eGFR is measured with error (whether due to variability of the underlying "true" GFR, error due to the misestimation of GFR by eGFR -- up to 35% CKD Stage misclassification using the CKD-Epi formula: Levey *et al.* (2009)[1], error due to the underlying determination of serum creatinine (Scr) which is increased at low levels of Scr because of its inverse relationship with eGFR, or variability in timing of the underlying measurements of serum creatinine). Mixed methods also permit specification of alternative models of change over time, permitting exploration of the hypotheses that progression is not linear. They would also facilitate development of time dependent models, allowing input of changes to patient's diabetes, hypertension, and cancer status over time, rather than limiting those variables to the baseline analysis. Most important, this sort of analysis might permit identification of the subset of individuals for whom the risk of progression is the highest -- that might be most benefited by careful follow-up of eGFR, and might give better guidance as to cost-effective testing intervals.

The authors' choice of analytic methods in this analysis should be put into perspective. The authors were evidently motivated to undertake this analysis so that cost-effectiveness studies could be done specifically in terms of the KDOQI/KDIGO CKD staging mechanism -- presumably to make it easier to develop easily followed guidelines for testing frequency. (See the authors' comment and reference 29 in their final Conclusion.) If the authors felt constrained by these requirements, they have performed a reasonable, if inefficient, analysis. But they and the authors of reference 29 might have done better to request and perform a more efficient analysis using the

underlying eGFR data rather than the derived CKD Stages to develop their model, and then abstracted stage specific recommendations from the more efficient model, rather than the other way around.

In sum, the analysis as performed using the specified analytic model was competently done, and the overall conclusions are justified - that the rate of progression of renal insufficiency is slow on average, and that dependence on this model would be associated with potentially significant error. But these limited conclusions don't exclude the possibility that more clinically reliable and useful data might be obtained by use of a method that didn't throw away so much of the available data.

References

1. Levey AS, Stevens LA, Schmid CH, Zhang YL, et al.: A new equation to estimate glomerular filtration rate.*Ann Intern Med*. 2009; **150** (9): 604-12 PubMed Abstract | Publisher Full Text

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

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

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

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: Nephrology, Epidemiology, Statistics, Clinical trial design, performance, and analysis.

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.

Author Response 27 Jul 2022

Benjamin Feakins, University of Oxford, Oxford, UK

Reviewer: The authors of this manuscript report on the long-term rate of progression of

chronic kidney disease (CKD) in a very large cohort of UK general practice patients followed as part of the Clinical Practice Research Datalink in the time between 1 April 2005 and 31 March 2014. Progression of kidney disease over each patient's course was assessed as the rate of progression from one stage to the next across the six currently standard KDOQI/KDIGO "stages" of renal insufficiency (Stages 1, 2, 3a, 3b, 4, and 5), which are defined by ranges of estimated glomerular filtration rate (eGFR). The rate of progression from one stage to the next was modeled as a Hidden Markov Model, which assumes that subjects are at "true" eGFR stages at each evaluation point, but that these stages may be misclassified based on random error of the eGFR estimate, so that the "observed" transitions may be misidentified. The overall conclusions of the authors are: a) that progression of renal disease to renal failure occurs quite slowly in general, with only a small fraction of patients that progress to kidney failure, and b) that progression of kidney disease as evidenced by [exclusive] use of progression in the KDOQI/KDIGO Stage of CKD, though generally reflective of the average course of kidney disease in the community, is associated with significant error when applied to an individual. Though the authors don't say this, the implication is that use of their analysis may not be very helpful in establishing guidelines as to in whom, and at what repeated intervals, eGFR determinations would be cost effective.

Authors: We disagree. We believe our model (unlike others that assume the observed eGFR is equal to true eGFR) permits analysis that correctly considers the misclassification that commonly occurs when monitoring kidney function using estimated GFR. This is important because when a person is misclassified into a higher stage treatment may represent a cost but provide less benefit than expected and in the opposite scenario a misclassification into a lower stage (false negative) would not incur costs from direct treatment but would render the patient at a higher risk for longer. Attempts to model the cost-effectiveness of monitoring without taking this into account are potentially misleading. The work presented in this manuscript formed part of a wider research project, which is intended to assess the cost effectiveness of kidney function monitoring. The health economics study using the outputs from this paper can be found at https://doi.org/10.1371/journal.pmed.1003478. The Health and Technology Assessment programme this work feeds into can be found at https://www.journalslibrary.nihr.ac.uk/programmes/pgfar/RP-PG-1210-12003.

Reviewer: The major strength of this study is the almost unique test bed offered by access to the data of almost 2,000,000 UK qualified general practice patients from Clinical Practice Research Datalink with "up-to-standard" data collection. Given the choice of analytic method, the authors' analysis is technically entirely competent. The assumptions and methods are well described. The results are clearly explained and put in appropriate context in the Discussion Section. The major limitation of the available data is that patients with the more advanced Stages of kidney disease are likely to have been referred to specialist providers, so that data on the more advanced stages of kidney disease, where progression is generally both more rapid and more clinically important, are more limited and more susceptible to bias. The authors clearly acknowledge this. Further, the issue addressed by this manuscript is potentially important, at least from a health economics point of view. Current policy recommendations uniformly encourage periodic assessment of renal function "in people with, or at risk of, chronic kidney disease (CKD)." But as noted above, in general kidney disease progresses quite slowly and only a small minority of subjects even

with documented kidney disease ever progress to chronic renal failure. There is almost no empirical data upon which to make any data-based recommendation about in whom or how often the "periodic assessments" should be done.

Authors: We would like that thank the Reviewer for the very positive comments and we agree that patients with more advanced stages of kidney disease might have been referred to specialist care. We would like to emphasise that the main aim of this study was estimate the rate of progression of kidney function decline before they enter end-stage renal disease, requiring specialist care. Therefore, there is a plethora of data, on which to make data-based recommendations regarding the frequency of periodic assessments. Tables S17–20 show that where such recommendations have been made, these have 'only' been for patients in eGFR stages G3a and G3b. Patients in these stages are still monitored within primary care, as opposed to specialist care. Moreover, as general practices are financially incentivised by the Quality and Outcomes Framework (QOF) to monitor such patients, there is no shortage of data for these patients as demonstrated in Table 1. Furthermore, it's also worth noting that Table 1 represents the baseline values for patients from an open cohort spanning the years 2005–2014. During this time, there were two substantial changes to QOF that increased the frequency of serum creatinine and proteinuria testing—one in 2006 and another in 2009. The effect of these changes can be seen in Figure 3. This figure shows that serum creatinine testing increased in subjects with CKD stages $\geq$2 from 2005 to 2008, where it reached a new plateau, and proteinuria testing increased in subjects with CKD stages $\geq$4 from 2008 to 2010, where it too reached a new plateau. A significant number of patients would have entered our study cohort prior to these dates, making Table 1 a poor reflection of the general quality of the testing data https://doi.org/10.1136/bmjopen-2018-028062. This figure shows that serum creatinine testing increased in subjects with CKD stages $\geq$ 2 from 2005 to 2008, where it reached a new plateau, and proteinuria testing increased in subjects with CKD stages $\geq$ 4 from 2008 to 2010, where it too reached a new plateau. A significant number of patients would have entered our study cohort prior to these dates, making Table 1 a poor reflection of the general quality of the testing data.

Reviewer: But the choice of analytic method is subject to criticism. The continuous variable eGFR may range from over 100 (mL/min/1.73 m^2) to close to 0. The reduction of this range of values to a classification with 6 levels is associated with massive information loss. Markov models deal with chances of transition from one state to another, and all members of one state are considered to be interchangeable, so that a patient with an eGFR of 100 (at Stage 1/2 with the combination of these two stages) is treated as equivalent to one with an eGFR of 62. But surely the chance of progression from an eGFR of 100 to an eGFR < 60 (Stage 3a) is much lower than the chances of progression from a value of 62 to < 60. Similarly, it is intuitively likely (though not at all certain) that a subject with a history of rapid loss of eGFR in the past will have more rapid progression in the future. But with a Markov model, all history is lost, and only the current stage, and not the prior history is taken into account. Given the huge amount of information available for this analysis, the inefficiency of the hidden Markov model may be overcome in the analysis of the overall community estimates of rate of progression. But these results will be of almost no use in prediction of the progression of kidney disease in the individual patient. And it is the individual patient's progress, not the average rate of progression in the community, that we are trying to clarify in recommending individual testing eGFR and setting repeat testing intervals.

Authors: The Reviewer is correct in stating that the Markov assumption means that each subject's transition to a more severe eGFR stage depends only upon their current stage. We agree that this assumption may be an oversimplification of reality, as it would be likely that an individual who has progressed rapidly through previous stages will continue to do so, and vice versa. As for the reviewers point about likelihood of progression being different if eGFR was 100 vs 62 ml/min/1.73m^2, we do not disagree that progression to a next stage would be different, but we would emphasise the problem of knowing what true eGFR is. For example, someone in whom true eGFR is 80 ml/min/1.73m^2, their observed eGFR could be anywhere between 70 and 90 ml/min/1.73m^2. So, by categorising eGFR into six levels there is some loss in information, but we have a more robust inference. We strongly disagree that these models will be almost of no use. Our models provide average progression times conditional on eGFR stage and presence of albuminuria, and other co-morbidities, e.g., heart failure. The same problem would exist for mixed effect models in as much as you can only provide aggregated rates of decline. We have added the following paragraph to the Discussion section:

"*We could be criticised for using an approach that categorises kidney function rather than a method that models continuous eGFR, such as generalised linear mixed models. This is because categorisation can lead to loss of information and reduced statistical power. However, such information loss is typically small when the number of categories is large [Altman D and Royston P, 2006], as was the case in our study. Furthermore, the use of categories that naturally aligned with clinically meaningful eGFR stages added to the interpretability of our findings. In addition, HMMs assume that all individuals within a state are interchangeable, and that the chance of progression to subsequent states depends only on the current state. This assumption may not hold if a patient whose kidney function has previously rapidly declined continues along this trajectory. To mitigate this, we have included updated risk factor information in our model and stratified by baseline albuminuria status. Further research could include assessing the impact of the Markov assumption on predicting kidney function decline using HMMs.*"

Reviewer: An alternative approach would have been to use a hierarchical ("mixed") analytic method (or perhaps even better, joint analysis combining the hierarchical analysis with a time to event analysis for ESRD and death) to evaluate the overall community rate of progression while estimating the distribution of individual rates of progression. Mixed models deal automatically with the problem that eGFR is measured with error (whether due to variability of the underlying "true" GFR, error due to the misestimation of GFR by eGFR -- up to 35% CKD Stage misclassification using the CKD-Epi formula: Levey *et al.* (2009)[1], error due to the underlying determination of serum creatinine (Scr) which is increased at low levels of Scr because of its inverse relationship with eGFR, or variability in timing of the underlying measurements of serum creatinine). Mixed methods also permit specification of alternative models of change over time, permitting exploration of the hypotheses that progression is not linear. They would also facilitate development of time dependent models, allowing input of changes to patient's diabetes, hypertension, and cancer status over time, rather than limiting those variables to the baseline analysis. Most important, this sort of analysis might permit identification of the subset of individuals for whom the risk of progression is the highest -- that might be most benefited by careful follow-up of eGFR and might give better guidance as to cost-effective testing intervals.

Authors: The reviewer is correct, there are alternative statistical models to the approach we led with in this paper. Our initial intention was to model eGFR (or log eGFR) using mixed effect models, but we decided against it for two reasons: the trajectories of eGFR are not linear or necessarily even monotonic in time; more flexible modelling is limited by the fact that for many people only a few eGFR measurements existed in their health record. The models that we did fit did not meet the assumptions of normality implicit in the random effects approach and this, we believe, led to poor predictive performance of these models. It is possible that in an alternative data set, it would be possible to fit these models, but this was not our experience using routinely collected GP data. We would like to stress that we used updated covariates in all the HMMs used in this study and did not condition on only the baseline values of these.

Reviewer: The authors' choice of analytic methods in this analysis should be put into perspective. The authors were evidently motivated to undertake this analysis so that cost-effectiveness studies could be done specifically in terms of the KDOQI/KDIGO CKD staging mechanism -- presumably to make it easier to develop easily followed guidelines for testing frequency. (See the authors' comment and reference 29 in their final Conclusion.) If the authors felt constrained by these requirements, they have performed a reasonable, if inefficient, analysis. But they and the authors of reference 29 might have done better to request and perform a more efficient analysis using the underlying eGFR data rather than the derived CKD Stages to develop their model, and then abstracted stage specific recommendations from the more efficient model, rather than the other way around.

Authors: The reviewer is not entirely correct. This model was one part of a larger funded project which included a cost-effectiveness study using the outputs from this model, but it is not true that we felt or were constrained by this. Our initial plan was to model eGFR as a continuous measure, but we did not for reasons explained in the previous response. We feel the objective of our work is quite the opposite to the reviewer and we would prioritise models than fit clinical practice over and above statistical efficiency.

Reviewer: In sum, the analysis as performed using the specified analytic model was competently done, and the overall conclusions are justified - that the rate of progression of renal insufficiency is slow on average, and that dependence on this model would be associated with potentially significant error. But these limited conclusions don't exclude the possibility that more clinically reliable and useful data might be obtained by use of a method that didn't throw away so much of the available data.

Authors: We disagree with the assertion made by the reviewer that our method “threw so much of the data away”. We agree that categorisation can result in loss of information, but this is most likely to occur when continuous measures are split in two (dichotomisation) and applied to small data sets. The thresholds for categorisation should also not be defined post-hoc. Neither of these are true in our case. To cite Altman and Royston “using multiple categories (to create an “ordinal” variable) is generally preferable to dichotomising. With four or five groups the loss of information can be quite small, but there are complexities in analysis” (see https://doi.org/10.1136/bmj.332.7549.1080). Moreover, we have provided calibration plots in Figures S1–4 that demonstrate the accuracy of the predictions made by

the models.

Competing Interests: No competing interests were disclosed.

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.

Author Response 27 Jul 2022

Benjamin Feakins, University of Oxford, Oxford, UK

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

Authors: We did not model or attempt to estimate the incidence of dialysis or transplantation. We have included patients who would be candidate for dialysis or transplantation (patients with stage 4/5 CKD), but we censored patients at the first code relating to dialysis or transplantation from hospital or GP records. Moreover, we excluded patients with history of renal replacement therapy. It is unlikely that a significant number of patients would have received either dialysis or renal replacement therapy without there being a code to identify it in the data.

Reviewer: 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.

Authors: It is true that kidney function, albeit measured by an imperfect test (e.g., eGFR) has been observed to improve and whilst this might be true, observed improvements in kidney function could also be as a result of measurement error or intervention. For simplicity, and with the agreement of the clinicians on the paper (DL and CO'C), we thought the unidirectional model for true underlying kidney function to be biologically plausible. We did perform a sensitivity analysis in which we allowed underlying kidney function to spontaneously (without intervention) improve. We concluded that this model did not fit the data as well as the progression only model. We have added the following text to the eighth paragraph of the Discussion section:

"*The assumption that true underlying kidney function only deteriorates with age is a fundamental part of the model and further research could investigate alternative models for underlying kidney progression and their impact monitoring recommendations.*"

Reviewer: 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).

Authors: This is a valid point, relating to a problem we have encountered in routine health record data sets. However, we feel that this issue is likely to have minimal impact to our study, as apart from the last 3.25 years, our data pre-dates the use of IDMS as the changes took place in 2010. Furthermore, we conducted this study in an open cohort (CRPD-Vision) that has been decreasing in size (https://doi.org/10.1136/bmjopen-2017-020738), further attenuating the pool of potentially affected tests.

Reviewer: 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.

Authors: In this study, we used all testing sources available to obtain eGFR, including pre-derived eGFR values. Unfortunately, the data source used provides no information on the techniques used to measure serum creatinine levels or to estimate eGFR. We also no longer have access to this data. Hence, we cannot say whether the calculation of a particular eGFR value was performed using laboratory methods or via an estimating equation in a GP practice. However, we estimated the overwhelming majority of eGFR values from raw serum creatinine values. Hence, any bias present from the use of pre-derived eGFR values is likely to be minimal. The method we used to calculate eGFR from serum creatinine was the CKD-EPI equation. We are aware that better method for the calculation of eGFR exist, however, the one we used would have been the one predominantly in use in UK clinical practice throughout the study period.

Reviewer: 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.

Authors: The requirement of three serum creatinine tests was imposed to establish a reliable baseline eGFR status (requiring two tests) and to estimate change over time. We concede that the population present in this sample is not representative of a general UK population so much as a UK general practice population, and we have amended the wording of the fourth paragraph of the Discussion section of the manuscript to reflect this.

This reads:

"*Inclusion into the study was conditional upon having three or more serum creatinine measurements, but creatinine is commonly measured in UK general and not necessarily for the purpose of monitoring kidney function or diagnosing kidney disease.*"

... and later in the eighth paragraph of the Discussion section:

"*Our study design means that we would also miss patients with ESRD who had not engaged in primary care.*"

The degree to which this could bias our results is open to some conjecture, and we have also made such comments in the Discussion section of the manuscript. Furthermore, others [Gruger J et al., 1991] who have studied the underlying mechanisms for requesting kidney function testing have concluded that the only mechanism by which testing may bias estimates of transition rates is that of patient-driven testing, which we feel is possible, but unlikely.

Reviewer: 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.

Authors: When we wrote the section of text in question, we were referring to the non-selected nature of any one individual appearing in the dataset, i.e., their presence was only determined through virtue of their general practice opting into the database. However, as per our previous response, the population used in this study is still representative of a GP patient cohort, which is the setting in which we anticipate these results to be generalised. We have endeavoured to ensure that our study population was as non-specific as possible, but he nature of our study meant that patients needed to have creatinine measures present in their medical history for us to draw any kind of inference on their kidney function decline.

We would also add that, contrary to the reviewer's comments, we have stratified our analysis by the factors most important in determining the rate of kidney function decline, being the initial albuminuria stage. This is in line with the findings of other studies in this area [Chang WX et al., 2016; Inaguma D et al., 2017; Coresh J et al., 2019].

Competing Interests: No competing interests were disclosed.

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