Public Health England

Protecting and improving the nation's health

Technical document for the chronic kidney disease prevalence model

About Public Health England

Public Health England exists to protect and improve the nation's health and wellbeing, and reduce health inequalities. It does this through advocacy, partnerships, world-class science, knowledge and intelligence, and the delivery of specialist public health services. PHE is an operationally autonomous executive agency of the Department of Health.

Public Health England
Wellington House
133-155 Waterloo Road
London SE1 8UG
Tel: 020 7654 8000
www.gov.uk/phe
Twitter: @PHE_uk
Facebook: www.facebook.com/PublicHealthEngland

Prepared by: Emma Barron
For queries relating to this document, please contact: ncvin@phe.org.uk

Published October 2014
PHE publications gateway number: 2014386

Introduction

The chronic kidney disease (CKD) prevalence model provides estimates of total (diagnosed and undiagnosed) CKD prevalence for adults aged 16 and over in England.

CKD is a sustained loss of kidney function and / or kidney damage. The most common consequence of CKD is premature mortality, particularly cardiovascular disease, but CKD can also progress to end stage renal disease (ESRD) with a need for treatment with dialysis and / or transplantation (renal replacement therapy (RRT)). CKD is often asymptomatic, frequently unrecognised and commonly occurs with a variety of other conditions such as hypertension and diabetes.

The model was developed using data from the Health Survey for England (HSE) – 2009 and 2010 and the 2011 Census. The estimates have been adjusted for age, sex, ethnicity, household tenure and sex-tenure interaction. As with all modelled data there is a degree of uncertainty around the estimates, therefore 95% credible intervals (CI) have been calculated that give a plausible range in which the true value is likely to be contained.

CKD estimates have been produced at local authority (LA) lower level, LA upper level, clinical commissioning group (CCG), region and the whole of England. The estimates are available to download at: www.ncvin.org.uk

Previous CKD modelling

CKD prevalence within England was first modelled using the New Opportunities for Renal Intervention by Computerised Assessment (NEOERICA) project in 2005. Individuals with CKD were identified using GP practice data from Kent and Salford. CKD prevalence was calculated by applying the estimated CKD prevalence by age grouping to the ONS 2007 population estimates. NEOERICA estimated the prevalence of CKD stage 3–5 in the English population aged 18 years to be 8.8%. These estimates however only take into account GP diagnosed CKD and the areas selected were not fully representative of the English population.

In 2011 East Midlands Public Health Observatory (EMPHO) and NHS Kidney Care produced an online toolkit that estimated CKD prevalence at Strategic Health Authority (SHA), LA and Primary Care Trust (PCT). Expected CKD prevalence was calculated using the combined results of the HSE 2009 and 2010. These estimates only adjust for age and sex and do not take into account other factors that may affect CKD prevalence such as deprivation and ethnicity.

Definitions

Current guidelines define moderate to severe CKD as either kidney damage or a glomerular filtration rate (GFR) less than 60ml/min per 1.73m^2 for three months or more, regardless of cause. The five different stages of severity of CKD are shown in Table 1 with stages 3–5 representing moderate to severe CKD.

Table 1: Different Stages of CKD

Stage	GFR (ml/min per 1.73m^2)	Description
1	≥90	Normal GFR and damage
2	60-89	Slight decrease in GFR and damage
3(a)	45-59	Moderate decrease in GFR
3(b)	30-44	
4	15-29	Severe decrease in GFR
5	<15	End stage renal disease

There are different methods used to evaluate GFR derived from measurement of serum creatinine, a breakdown of muscle protein. The Modification of Diet in Renal Disease (MDRD) equation has been the most widely used equation to estimate GFR and accounts for differences in patient age, sex and ethnicity, all of which influence production of serum creatinine. The Chronic Kidney Disease Epidemiology (CKDEPI) equation was developed in 2009 and from July 2014 was recommended in NICE guidelines. The CKDEPI equation was found to predict CKD as well as the MDRD equation in patients with moderate CKD and perform even better than the MDRD equation in predicting CKD stage in people with higher GFRs. However, the MDRD equation was used in these analyses since the estimates were produced before the updated NICE guidelines and MDRD is still in routine use in clinical laboratories. CKDEPI estimates have also been calculated and will be released when the CKDEPI equation becomes more widely used.

MDRD equation:
GFR = 175 x $S_{Cr}^{-1.154}$ x Age $^{-0.203}$ x (0.742 if female) x (1.21 if African American)

Where S_{Cr} = serum creatinine

Creatinine samples were assayed using an isotope dilution mass spectrometry traceable enzymatic assay in a single laboratory (Clinical Biochemistry Department at the Royal Victoria Infirmary (RVI), Newcastle-upon-Tyne) using the Roche Modular P analyser, with samples stored, frozen at -40^{o}C, and then thawed for measurement.

Data sources

The 2009 and 2010 Health Survey for England (HSE) were combined and used for analyses, totalling 17,040 people (8,620 from 2009 HSE and 8,420 from 2010 HSE). Participants in the survey were selected using a multistage stratified random probability sample. Therefore, participants in the survey were selected over two stages. The first stage involved selecting the higher level cluster, referred to as the Primary Sampling Unit (PSU). These PSUs were based on postcode sectors randomly selected within regions based on postcode sectors from the postcode address file (PAF); a file which contains all known postcodes within the UK. The second stage involved selecting the lower level cluster, referred to as the secondary sampling unit (SSU). These SSUs consist of a list of addresses and represent the household level. Households were randomly selected within each SSU.

For both 2009 and 2010 HSE, data collection was followed by a nurse visit which included collection of blood, saliva and urine samples. People under the age of 16 or without a valid serum creatinine value to determine CKD were excluded. This left a total of 6,046 people included in the analysis (2,199 from 2009 HSE and 3,847 from 2010 HSE).

Cross-tabulations of individual level local statistics from the 2011 Census were also used to provide counts of the number of individuals in each small area who fall into particular socio-demographic categories. A decision was made to cross-tabulate by age, sex and ethnicity.

Model construction

Only variables that were present in both the HSE and 2011 Census simultaneously were included in the modelling framework. A simple regression-based CKD prevalence model was not appropriate as the source data, the HSE, is a cluster randomised survey and analysis should reflect design. Modelling individual associations with CKD is important but Census constraints limit the number of individual variables that can be modelled. Additional variables can be included to reflect the characteristics of the areas where HSE respondents live. Therefore a multilevel small area synthetic estimation (ML-SASE) methodology was used.

ML-SASE modelling is an extension of linear regression in which an equation estimates a target response variable in terms of a number of predictor variables. Predictor variables within a multilevel modelling framework relate to either individual level influences or area level responses and these two influences can interact. Individual level variables included in the modelling were age (16-34, 35-54, 55-64, 65-74 and 75+) sex (male, female) and ethnicity (white, south Asian, black, other) and area level variables included were tenure (owned, rent), vehicle ownership (at least one, none), general health (great, poor) and limiting long term illness (LLTI) (yes, no). The response variable was defined dichotomously as if an individual had CKD (GFR <60ml/min per 1.73m^2) or not. Note, CKD was determined by a single blood test and not two blood tests three months apart as recommended in NICE guidelines and therefore may represent an overestimate of population prevalence.

Variables were retained in the final model if they were found to be statistically significant. Tests for interactions between age, sex and ethnicity terms were also carried out and were retained if they were found significant. Tests from cross-level interactions between individual and area characteristics were also carried out, with any significant variables retained. All tests of significance were carried out using the Wald test (z-test). A decision was made to retain at least one ethnicity variable as non-SASE models showed the importance of ethnic associations with CKD.

The model was computed using MLwiN (version 2.24, Centre for Multilevel Modelling, University of Bristol). Level 1 was defined by the serial number of each individual and level 2 was defined as the PSU cluster that each individual lives within. Initial multilevel models were produced using iterative generalised least squares with first order maximum quasi-likelihood estimation. Once a final model was selected, a Monte Carlo Markov Chain (MCMC) approach was used to refine the model and allow for more robust estimates and standard errors. The MCMC process applied a burn-in of 50,000 iterations.

The Model

Significant variables retained in the model were age, sex, Asian ethnicity, tenure and a tenure–sex interaction. Logit values from the final model were untransformed to get the proportion of each age-sex-ethnic group with CKD (40 in total) at middle super output area (MSOA) with CKD. The age/sex/ethnic proportions of CKD were then adjusted for tenure and tenure-sex interactions, applied to age-sex-ethnic counts from 2011 census data at MSOA level and aggregated to find the total number of CKD cases for each MSOA. These estimates were then aggregated to the geographical areas of interest (LA lower level, LA upper level, CCG and Region) and divided by the 2011 ONS population estimates for that geographical area to get the population prevalence of CKD for each geographical area.

The final model variables, with their logit and standard error values, are shown in Table 2 and the univariate odds ratios (ORs) of having CKD with 95% confidence intervals are shown in Table 3. The results in Table 2 indicate that the chance of having CKD is approximately 0.1% if you are categorised as the stereotypical individual (male, aged 16-35, and living in a typically average area in terms of ethnicity and tenure, and equate to all of the listed interactions). This percentage is obtained by taking the antilogit of -6.833. However, if you happen to be aged 75+, but otherwise remain the stereotypical individual, the chance of having CKD increases to approximately 28%.

Table 2: Logit terms and standard error

	Logit	Standard Error
Constant	-6.833	0.614
Sex: female	0.361	0.116
Age: 35-54	2.801	0.625
Age: 55-64	3.795	0.624
Age: 65-74	4.780	0.618
Age: 75+	5.900	0.616
Ethnicity: non-Asian	0.396	0.733
Tenure: Rent	1.000	0.486
Rent: Female	-1.646	0.587

The ORs in Table 3 imply increased risk of CKD 3-5 in older people as expected and an increased risk of CKD in women, which partly reflects use of the MDRD equation. People who rent are at an increased risk of CKD, highlighting an expected link with socio-economic status. The tenure–sex interaction implies that the impact of rental tenure is attenuated in women.

Table 3: Univariate odds ratios with 95% confidence intervals

Variable		Odds ratio	Confidence intervals
Sex	Male	1	
	Female	**1.54**	**1.24 – 1.93****
Ethnicity†	White	1	
	South Asian	**0.06**	**0.09 – 0.42****
	Other	**0.15**	**0.03 – 0.77****
Age	16-34	1	
	35-44	**13.75**	**4.03 – 46.95****
	55-64	**37.77**	**11.09 – 128.58****
	65-74	**100.06**	**29.78 – 336.20****
	75+	**321.83**	**96.60 – 1072.18****
Tenure	Own	1	
	Rent	**1.38**	**1.08 – 1.77***
Tenure*Sex	Male: Own	1	
	Female: Own	**1.54**	**1.19 – 1.99****
	Male: Rent	1.44	0.96 – 2.16
	Female: Rent	1.12	0.80 – 1.56

†No Black individuals were classified as having CKD for 2009-10 HSE data
*p<0.05 **p<0.01

Model validation

A number of different tests have been used in order to internally validate the CKD prevalence models.

1. Assessment of fit

    1.1 Area level variance explained by ML-SASE model

It is recommended that a minimum of 40% of area variance is explained by the final multilevel model compared to the null model measured by the variance partition coefficient (VPC). The percentage of area level variance explained by the model was 66.1%, therefore meeting this requirement.

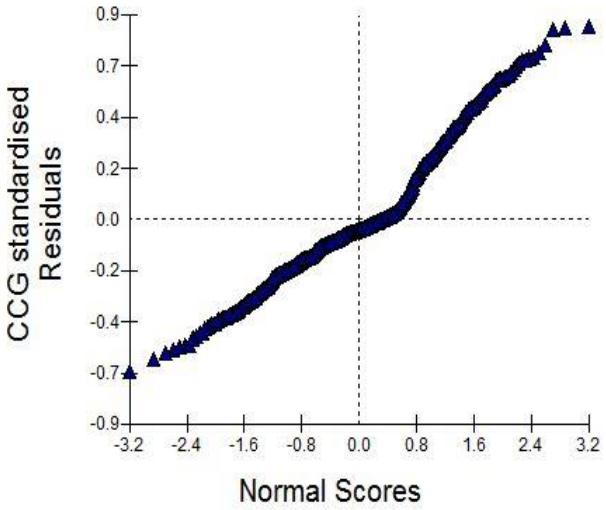
    1.2 Standardised residuals against ranked residuals

Normality plots of standardised residuals against ranked residuals were used to determine if the assumption of normality was valid; straight lines indicating that this was met (Graph 1). The line is reasonably straight, therefore this assumption was upheld.

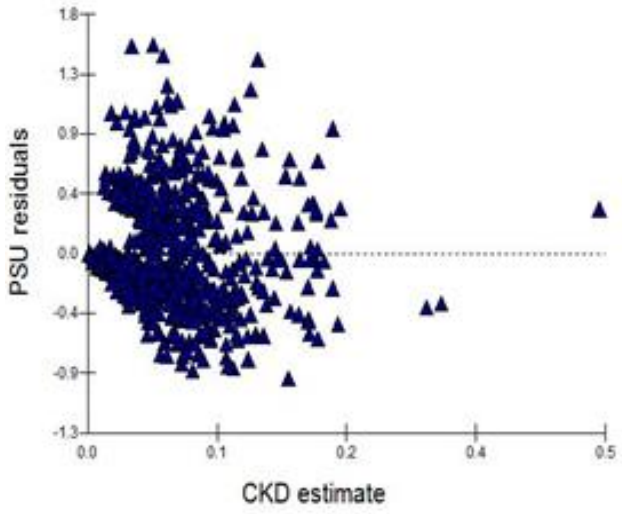
1.3 Standardised residuals against predicted CKD proportions

Analysis of the model residuals can be used to assess model fit by plotting the standardised residuals against the predicted values. If the model is a good fit there should be no relationship between the estimates and residuals. The plot of the standardised residuals against the predicted CKD proportions at the PSU level shows no relationship, therefore this was a good model fit (Graph 2).

Graph 1: Normality plots of standardised residuals against ranked residuals



Graph 2: Standardised residuals against CKD estimates



2. Percentage correct prediction

The percentage correct prediction (PCP) is used to see how good the model is at predicting the original survey responses. For an event with even probability of occurring and not occurring, the PCP assumes that if the estimated proportion is greater than or equal to 0.5 then the event is expected to occur, if p is less than 0.5 then the event is expected not to occur. For the CKD model, p was set to 0.0610 (mean CKD probability). The percentage correctly predicted was 79.7%.

In addition to assessing the robustness of the modelling process, a complete evaluation of the ML-SASE estimates must also recognise that they depend on two data sources which had a differential completion of survey data questions and a level of non-response in the Census.

External validation

In order to externally validate the ML-SASE model estimates, they were compared to the EMPHO estimates (which applied HSE 2009 and 2010 CKD prevalence rates to age and sex specific populations) and the 2011/12 Quality and Outcomes Framework (QOF) CKD disease register (a financially rewarding voluntary incentive scheme for GP practices in the UK).

Compared to the EMPHO estimates, the ML-SASE model estimates at CCG level were nearly all 0-1% lower. Only three CCGs had differences of more than 1%: East Lancashire (1.2%), North Manchester (-1.1%) and Norwich (-1.3%). The ML-SASE estimates also correlate very strongly with the EMPHO estimates; the correlation coefficient was 0.993 ($p<0.001$).

Compared to QOF estimates, the ML-SASE model estimates CKD prevalence for the vast majority of CCGs to be 2% greater.

Calculation of credible intervals

It is not recommended to calculate confidence intervals for the model estimates since they are synthetic estimates and therefore a biased estimate of the true value. Monte Carlo Markov Chain (MCMC) estimation was used as an alternative and was used to calculate 95% credible intervals around the prevalence estimates.

MCMC is an iterative procedure that generates a series of values for each parameter. This approach refines the model and allows for more robust estimates and standard errors. The sample of values generated for each parameter was used to produce credible intervals for the estimates. In order to ensure consistency and to keep within IT processing power limitations, the number of MCMC iterations differed for the different levels of geography; LA lower level – 25,000 MCMC iterations, LA upper level – 25,000 MCMC iterations, CCG – 38,000 MCMC iterations; and Region – 100,000 MCMC iterations.

CKD prevalence projections

The future prevalence of CKD was calculated by applying estimated CKD prevalence by age groupings (16-34, 35-54, 55-64, 65-74 and 75+) to the CCG 2012 based subnational population projections (SNPP) produced by the Office for National Statistics (ONS). The estimates did not adjust for other socioeconomic factors (such as tenure), assumed no change in the age-specific prevalence of CKD, and assumed no improvement in the prevention and management of CKD in the population.

Contact:

The CKD model was developed by: Grant Aitken (PhD Researcher, Geography and Environment, University of Southampton) supervised by Graham Moon (Professor in Spatial Analysis, Geography and Environment, University of Southampton) and Paul Roderick (Professor in Public Health, Primary Care and Population Sciences, University of Southampton).

Email the National Cardiovascular Network (NCVIN) for further details: ncvin@phe.gov.uk