ABSTRACT

With the introduction of the National Kidney Foundation Kidney Disease Outcomes Quality Initiative chronic kidney disease (CKD) guidelines, CKD has been identified as common, particularly in the elderly. The outcomes for those with CKD can be poor: mortality, initiation of renal replacement therapy, and progressive deterioration in kidney function, with its associated complications. In young people with CKD, the risk of poor outcome is high and the social cost substantial, but the actual number of patients affected is relatively small. In the elderly, the risk of poor outcome is substantially lower, but due to the high prevalence of CKD the actual number of poor outcomes attributable to CKD is higher. Predicting which patients are at greatest risk, and being able to tailor care appropriately, has significant potential benefits. Risk prediction models in CKD are being developed and show promise but thus far have limitations. In this review we describe the pathway for developing and evaluating risk prediction tools, and consider what models we have for CKD prediction and where next.

Keywords: Chronic kidney disease, outcome, risk prediction.

CHRONIC KIDNEY DISEASE: THE BURDEN OF CARE

The recognition of chronic kidney disease (CKD) prior to 2002 was inconsistent, with no standard definition and, in some cases, led to late referral of patients to specialised renal services. This was associated with poor outcomes on renal replacement therapy (RRT) and missed clinical opportunity to improve disease course. The introduction of the National Kidney Foundation Kidney Disease Outcomes Quality Initiative (KDOQI) CKD definitions, international acceptance of these definitions, and instigation of CKD guidelines has led to a recognition that the prevalence and incidence of CKD is far in excess of earlier estimates. Reports based on data from the National Health and Nutrition Examination Survey (NHANES) suggested 13.1% (26 million) of the US adult population in 2000 had CKD. The prevalence of CKD increases with age; in NHANES approximately 50% of those aged 70 years and older were reported to have CKD. This is set to grow as the global population ages. A recent study estimated that the lifetime risk of developing CKD Stage 3-5 was approximately 60%; three out of five babies born today are destined to develop CKD in their lifetime.

The natural history of CKD is perceived to be one of loss of renal function over time with associated complications, leading ultimately to the need for RRT and finally death. With declining estimated
glomerular filtration rate (eGFR) the risk of reaching end-stage renal disease (ESRD) increases, estimates vary. A Chronic Kidney Disease Prognosis Consortium (CKD-PC) patient-level meta-analysis reported increasing hazard ratios (HRs) with declining eGFR up to an adjusted ESRD HR of 51 (32-83) 95% confidence interval (CI) for eGFR<15 (versus 45-74 ml/min/1.73 m²).¹⁶ Worsening proteinuria is also associated with increasing risk of ESRD; CKD-PC report a HR of 9 (2-50) for ACR ≥1000 (versus <30) mg/g.¹⁶ All-cause mortality also increased with worsening eGFR and worsening proteinuria.¹⁷ Other outcomes for those with CKD include a high cardiovascular morbidity and mortality.¹⁸ Progressive deterioration in kidney function, short of the requirement for RRT, carries the risk of developing anaemia, acidosis, bone disease, a need to prepare for RRT, or instigation of conservative care. Recent CKD cohort reports confirm that, although many initiate RRT or die, a significant number may still be alive after several years. Our own data on Stage 3b-5 CKD suggest that approximately one-third are still alive and not requiring RRT at 5 years.¹⁹

CKD and the associated increased use of hospital services²⁰ come with a significant financial cost – estimated for England’s NHS during 2009-10 to be at £1.45 billion (~1.3% of NHS spending), of which just over half was due to RRT costs²¹ Instigation of RRT also involves other costs; these include personal loss of earnings, health, and personal-social costs both for patients and their carers. Thus, in the general population there is a high burden of CKD both in terms of volume of disease and cost of treatment and management.

**INDIVIDUAL RISK VERSUS POPULATION BURDEN: AGE MATTERS**

The effect of this difference in risk is illustrated in **Figure 1**. Relative risk defines the strength of the association between the outcome and the exposure (in this case CKD and RRT or death). Absolute measures of risk estimate the impact of a disease (here, CKD) on an individual, or indeed population. For an individual this would be the difference in risk of a given outcome depending on whether a person has the disease (CKD) or not - the excess risk associated with having CKD.²² For populations, the population attributable risk describes how much of an outcome can be accounted for by exposure to a particular disease in the population, and therefore takes into account the disease prevalence.²²

**Figure 1a** illustrates the prevalence of CKD for a population. The mortality in those without and with CKD are illustrated in **Figure 1b**; this uses figures from the CKD-PC meta-analysis reported by Hallan et al.²³ using the mean mortality rates for those with eGFRs of 80 and 45 ml/min/1.73 m², respectively to represent an average individual with ‘no CKD’ and ‘CKD’, respectively. For those aged 18-54 years the mean mortality rates were 4.0 and 13.0 per 1000 patient-years (py) for ‘no CKD’ and ‘CKD’ individuals, respectively, thus, individuals with CKD had an excess mortality risk of 9 per 1000 py, and a mortality HR or relative risk of ~3 as a result of CKD. Given that the prevalence of CKD is low in those aged 18-54, this excess personal risk translates into a relatively small excess number of deaths (the numbers shown in **Figure 1d** are purely illustrative since they are estimated using outcomes for those with an eGFR of 45 ml/min/1.73 m² to represent those with CKD, whereas the majority of those with CKD will have a far better eGFR than this). As shown in **Figure 1b**, for those aged ≥75 years, mean mortality rates for ‘no CKD’ and ‘CKD’ were 57.8 and 85.0 per 1,000 py, an excess individual risk of 27.2 per 1,000 py and a much lower relative risk of ~1.3. However, as illustrated in **Figure 1d** the far higher prevalence of CKD means that the excess deaths are far higher despite this lower relative risk.

The equivalent mean rates for achieving ESRD reported by Hallan et al.²³ in those with ‘no CKD’ and ‘CKD’ (eGFRs of 80 and 45 ml/min/1.73 m², respectively) by age are shown in **Figure 1c**. The rates were higher in the youngest groups, the relative risk was higher in the young also. The excess individual risks were 45.1 and 8.0 per 1,000 py for the youngest and oldest group, respectively.
However, since the numbers at risk are so small in the young, and large in the elderly, the excess cases of ESRD in the illustration are actually a little higher for those over 65 years of age. Thus, amongst the population there are two distinct groups – a group where CKD is uncommon but, for those with CKD the personal risk of poor outcomes are high, and a group where CKD is common but personal risk of poor outcomes is low. However, as a result of the number of individuals in this second group, they contribute a significant number of poor outcomes.

For care planning, although in terms of personal risk there are certain groups at high-risk, the majority of those that actually contribute to the highest volume of care requirements are generally older and have a lower personal risk. Although to have a specialist review of the younger high-risk group is not challenging, identifying those at risk in the low relative-risk elderly group might be, but it is important for care-planning. An ability to accurately determine which of the high volume group with low personal risk are at more risk than their peers would allow directed care for RRT planning, pre-dialysis care, specialist nephrology input, and mortality risk reduction steps. Also, very importantly, it would identify those who will not suffer these poor outcomes, for whom nephrology care is unlikely to be necessary. In other medical fields with high volume disease, such as ischaemic heart disease\textsuperscript{25,26} prediction models for the assessment of risk have been introduced and are commonly used in clinical practice.

**CKD OUTCOME PREDICTION AND CARE PATHWAY DEVELOPMENT**

Developing safe and effective prediction models takes time and is a complex process.\textsuperscript{27-32} Models are developed using previous experience and work of others to inform and check likely prognosis and prognostic variables in a development cohort. Regression models are often then used to relate these variables and the outcome; importantly, indices of model performance are measured to decide on the ‘best’ model. A description of how well the model performs demonstrates internal validity. To demonstrate that the model performs well in another similar situation, external validation in another population is needed. Once good performance has been demonstrated, ideally a randomised controlled trial (with economic evaluation) would be used to demonstrate the implications of introducing the model into clinical practice. Then finally, if the model improved care or outcomes and had acceptable utility, it would be implemented in clinical practice. However, many of these steps are often not done and many developed models remain untested in clinical practice.
Figure 1: Population chronic kidney disease (CKD) prevalence, personal risk, relative risk, and population attributable risk.

1a shows the prevalence of CKD in an illustrative adult population (based on unpublished data from Grampian Laboratory Outcomes Morbidity and Mortality 2 Study24); in the darker colour for both males and females are the number with an eGFR of <60 ml/min/1.73 m², percentage values shown refer to the age-bands as are demonstrated in b, c, d, and e. ** 1b, 1c, 1d, and 1e are based on rates published in Hallan et al.23 with those with ‘not CKD’ and ‘CKD’ are based on figures for those with eGFR 80 and 45 ml/min/1.73 m², respectively, although it should be kept in mind that the majority of those with CKD will have an eGFR>45 so this will overestimate CKD excess risk dramatically, but it is used here for illustration purposes only and applied to the population in 1a. 1b shows in light grey the death rates by age for those with an eGFR of 80 ml/min/1.73 m² as representative of ‘Not CKD’, in dark grey is shown the additional risk in those with an eGFR of 45 ml/min/1.73 m² as representing ‘CKD’, the * crude death rate ratio (based on the ratio of these two death rates) is shown in brackets. 1c shows similar for the occurrence of ESRD. 1d shows the excess number of deaths that could be expected in this population based on the difference between the death rates in those with and without CKD and the numbers with ‘CKD’ in the population. 1e shows the same for ESRD. Both are overestimates as the result of using the rates for 80 and 45 ml/min/1.73 m² as estimates for those with ‘not CKD’ and ‘CKD’.

In the field of renal medicine there are no prediction models for CKD outcomes that are in common use in clinical practice. But early work33 has been done that suggests prediction models could offer ways to plan and tailor care for people with CKD based on their risk of poor outcomes. Potential models have been developed to predict CKD outcome including mortality, cardiovascular disease, and kidney failure; the studies identified in two recent reviews34,35 are summarised in Table 1.

The identified prediction studies have shown that it is possible to develop models to predict outcome in those with CKD, using a wide range of data sources. The majority of these studies and models sought to predict ‘renal failure’ (most but not all defined by the initiation of RRT). Five of the identified studies report on models to predict mortality in those with CKD,34 sometimes as a composite end-point ‘death or renal failure’. Just three report cardiovascular event prediction...
models, including a comparison with the performance of the Framingham risk equation. These death and cardiovascular models rarely report most performance metrics, except discrimination which were usually reasonable (C-statistics of 0.60 to 0.82).34

Of the ten studies developing prediction models for renal failure, two studies were in individuals with IgA nephropathy alone, and two in people with diabetic nephropathy, limiting generalisability for those with CKD of other or unknown cause. Only two studies were in community populations not identified through referral to specialist nephrology care. The variables in the models were important in terms of application to current clinical practice: those that are routinely performed such as creatinine (and eGFR) are easily translated into care; using variables that are only measured if clinically indicated potentially limits utility as does the use of biological measures that are not in common use in routine clinical practice such as cystatin C and NT-pro-BNP. No studies used the pattern or rate of eGFR or creatinine change as a predictor variable.

<table>
<thead>
<tr>
<th>Study</th>
<th>Setting, included individuals</th>
<th>Size</th>
<th>Routine plus special measures</th>
<th>Stage in development</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimitrov36</td>
<td>Clinic, RCT, Nondiabetic CKD</td>
<td>344</td>
<td>Routine, calcium phosphate product</td>
<td>P. metrics (D+ C+) Internal validation</td>
</tr>
<tr>
<td>Keane37</td>
<td>Clinic, RCT, Diabetic nephropathy</td>
<td>1,513</td>
<td>Routine</td>
<td>P. metrics (D- C-) Internal validation</td>
</tr>
<tr>
<td>Wakai38</td>
<td>Clinic, cohort, IgA nephropathy</td>
<td>2,269</td>
<td>Routine, BP, haematuria Histological grade</td>
<td>P. metrics (D+ C-) Internal validation</td>
</tr>
<tr>
<td>Kent39</td>
<td>Clinic, RCTs, Nondiabetic CKD</td>
<td>1,860</td>
<td>Routine, BP</td>
<td>P. metrics (D+ C+) Internal validation</td>
</tr>
<tr>
<td>Johnson40</td>
<td>HMO cohort CKD</td>
<td>9,782</td>
<td>Routine, BP, history of DM</td>
<td>P. metrics (D+ C+) Internal validation</td>
</tr>
<tr>
<td>Goto41</td>
<td>Clinic, cohort, IgA nephropathy only with eGFR &gt;60 ml/min/1.73m²</td>
<td>790</td>
<td>Routine, BP, haematuria Histological grade</td>
<td>P. metrics (D+ C+) Internal validation</td>
</tr>
<tr>
<td>Hallan42</td>
<td>Population, cohort CKD and no CKD</td>
<td>65,589 (not all CKD)</td>
<td>Routine, BP, meds, DM, chol Physical activity</td>
<td>P. metrics (D+ C-) Internal validation Explores potential clinical impact</td>
</tr>
<tr>
<td>Landray43</td>
<td>Clinic, cohort, CKD</td>
<td>382</td>
<td>Routine, phosphate</td>
<td>P. metrics (D+ C+) Internal validation, External validation (213)</td>
</tr>
<tr>
<td>Desai44</td>
<td>Clinic, RCT Diabetics with CKD and anaemia</td>
<td>995</td>
<td>Routine, race, BMI, meds, Hx PAD/stroke/ HF/arrhythmia/AKI, ferritin, CRP TnT, NT-pro-BNP</td>
<td>P. metrics (D+ C-) Internal validation</td>
</tr>
<tr>
<td>Tangri33</td>
<td>Clinic, cohort, CKD</td>
<td>3,449</td>
<td>Routine, (8 variable model) phosphate, bicarbonate, calcium</td>
<td>P. metrics (D+ C+) Internal validation, External validation (4942)</td>
</tr>
</tbody>
</table>

**Cardiovascular events**

<table>
<thead>
<tr>
<th>Study</th>
<th>Setting, included individuals</th>
<th>Size</th>
<th>Routine plus special measures</th>
<th>Stage in development</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shlipak45</td>
<td>Population, cohort, CKD</td>
<td>1,249</td>
<td>Routine, race, education, meds, Hx of DM/ CVD, education, BP, BMI</td>
<td>P. metrics (D+ C-)</td>
</tr>
<tr>
<td>Weiner46</td>
<td>Population, cohort, CKD</td>
<td>934</td>
<td>Routine, BP, chol, DM, smoking, race</td>
<td>External validation of Framingham risk equation P. metrics (D+ C+) of recalibrated model</td>
</tr>
<tr>
<td>McMurray47</td>
<td>Clinic, RCT, Diabetics with CKD and anaemia</td>
<td>955</td>
<td>Routine, race, BMI, meds, Hx PAD/stroke/ HF/arrhythmia/ AKI, ferritin, CRP ECG TnT, NT-pro-BNP</td>
<td>P. metrics (D+ C-)</td>
</tr>
</tbody>
</table>
Routine might include: age, sex, some creatinine based measure of excretory renal function, some measure of albuminuria, some measure of serum protein, some measure of anaemia; CKD: chronic kidney disease; ESRD: end-stage renal disease; RRT: renal replacement therapy; HMO: health maintenance organisation; RCT: randomised controlled trial; BP: blood pressure or hypertension; DM: diabetes mellitus; CVD: cardiovascular disease; Hx: history of; PAD: peripheral arterial disease; HF: heart failure; BMI: body mass index; AKI: acute kidney injury; CRP: C reactive protein; TnT: troponin T; NT-pro-BNP: N terminal prohormone brain natriuretic peptide. P. metrics: performance metrics; D+/D-: discrimination reported/not reported (usually as ROC or C statistic); C+/C-: calibration reported/not reported (usually as Hosmer-Lemeshow statistic or plot); Chol: cholesterol; ECG; electrocardiogram; BMI: body mass index.

Reporting of the metrics considered of importance in the development of a prediction model (discrimination, calibration, model fit, and reclassification), summarised in Table 2 and previously was variable. Discrimination was reported in the majority and was excellent; c-statistics of 0.79-0.94 (compared to the widely accepted Framingham model performance of 0.75-0.81). Calibration was not consistently reported. The clinical implications were explored and reported in only one study, although model application to clinical care was discussed by two authors and useable interfaces presented by three. Only two authors demonstrated external validity. Both studies’ models were developed and externally validated in individuals referred to renal services and, as such, their performance in individuals from less specialised settings (e.g. community or non-specialist care) have not been demonstrated.

The best reported of these models were by Tangri et al., who recounted model calibration, and for the several models developed report the relative performance of one model over another. The original study included external validation and they have since been externally validated by others. The performance of the Tangri 3 and 4 variable models are illustrated in Figure 2 applying the models to an example CKD population cohort generated from our routine clinical practice in the North East of Scotland. The 3 variable model could be applied to the whole cohort; however, all 1,246 with Stage 4 CKD were labelled as high-risk when in fact only 89 had started RRT by 5 years (very sensitive, very poor specificity). Of the 4,951 with Stage 3b CKD (41 initiating RRT by 5 years) 2,053 were defined as high-risk (sensitivity 0.95, but specificity of only 0.59). The 4 variable model restricted the number for whom there was appropriate data available as part of routine care, so that only 13% of those with Stage 3b CKD could have a risk calculated. However, performance was a little better. Nevertheless, even the best performing of the Tangri models showed better discrimination amongst those with Stage 4 CKD than Stage 3 CKD where the main clinical challenge remains.

Table 1 continued.
Thus, the models currently available, although demonstrated to be useable, are limited by a number of issues and flaws and so need further work. Reporting and testing of model performance was inconsistent, and while standard, basic tests of internal validity of the models tended to show promise, more transparent testing on external validation cohorts reporting sensitivity, specificity, false-positives, and false-negatives would be helpful in assessment of clinical utility, particularly using the suggested thresholds for identifying high-risk individuals.

Models that are practical to use in ‘real-life’ with real-life data such as the Tangri model should be given priority, and model refinement should ensure practical real-life usability. Models need to be judged in terms of the added value for care and service planning over the current use of referral guidelines based on eGFR and proteinuria.53 The addition of novel biological markers has, thus

---

### Table 2: Common and expected indices for the reporting of prediction model performance.

<table>
<thead>
<tr>
<th>Description</th>
<th>How usually measured/reported</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Discrimination</strong></td>
<td></td>
</tr>
<tr>
<td>Whether the model assigns a higher probability to individuals who have the outcome of interest versus those who do not.</td>
<td>Area under the receiver operating characteristic (ROC) curve, or c-statistic. 0.50 is no better than chance. 0.70-0.80 good. &gt;0.80 excellent.</td>
</tr>
<tr>
<td><strong>Calibration</strong></td>
<td></td>
</tr>
<tr>
<td>How well the predictions equate to actual outcome.</td>
<td>Calibration plots; variants on the Hosmer-Lemeshow $\chi^2$ statistic, rank into deciles of predicted risk, then compare the expected and observed frequency of the outcome in each decile. If very different then not good calibration.</td>
</tr>
<tr>
<td><strong>Goodness of fit</strong></td>
<td></td>
</tr>
<tr>
<td>How well the model fits the data concerned.</td>
<td>Akaike or Bayes information criterion, the lower the value the better fit of the model.</td>
</tr>
<tr>
<td><strong>Reclassification</strong></td>
<td></td>
</tr>
<tr>
<td>How a newer (or more complex) model improves the reclassification of individuals - those with the event to a higher predicted risk and those without the event to a lower predicted risk.</td>
<td>Net reclassification improvement or integrated discrimination improvement. If the direction of the reclassification is correct - those with events to higher risk, those with no events to lower risk, then the newer (or more complex) model is better.</td>
</tr>
</tbody>
</table>

---

**In our local cohort with CKD, using an RRT risk of >5% at 5 years as high-risk, the Tangri 3 and 4 variable models performed as below:**

<table>
<thead>
<tr>
<th>CKD stage</th>
<th>3 variable model</th>
<th>4 variable model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients with CKD</td>
<td>4,951</td>
<td>1,246</td>
</tr>
<tr>
<td>Number start RRT at 5 years</td>
<td>41</td>
<td>89</td>
</tr>
<tr>
<td>Missed RRT cases</td>
<td>2 (5%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Number high-risk</td>
<td>2053</td>
<td>1246</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.95</td>
<td>1.00</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.59</td>
<td>0.00</td>
</tr>
</tbody>
</table>

**Figure 2:** Practical example of prediction model performance not illustrated using discrimination, calibration, and model fit indices.

CKD: chronic kidney disease; RRT: renal replacement therapy.
far, provided limited improvement in model performance and has restricted immediate clinical implications if testing for such markers is not in widespread clinical use. Greater gains for model refinement might come from taking into account the type of information used in clinical practice to assess long term risks: rate and pattern of kidney function decline, comorbidity, and underlying renal pathology, for example.

Model performance has not been explored in depth in older age populations. Experience from the cardiovascular literature suggests that model performance may not be as good in the elderly. In this age group, issues of competing risks from other morbidities and death become increasingly important. Further external validation of refined models in community settings such as primary care is required prior to any use in this context. Once these models have been refined, good quality randomised controlled trials of their introduction should be run to demonstrate any improvement in care and outcome delivery and absence of harm. This would then facilitate the introduction of stratified renal medicine appropriate to the risk profile of individuals concerned. Referral and intensive management with optimal implementation of current treatment guidelines could then be focused to those with the greatest opportunity to benefit. Health economic evaluation of the cost-effectiveness of different models of care delivery, based on stratification using prediction models, will be needed to support service planning, but the opportunity to reduce the need for referral to specialist services and frequency of follow-up has significant potential benefits for patients and health services.

**CONCLUSION**

With the introduction of the KDOQI CKD guidelines, CKD is being identified more commonly, particularly in the elderly where milder renal impairment is predominant. The outcomes for those with CKD are poor – mortality, initiation of RRT, and progressive deterioration in kidney function, with its associated complications. In young CKD patients, the risk of poor outcome is high and the social cost substantial, but the actual numbers of patients affected is relatively small. In the elderly, the risk of a poor outcome is substantially lower, but due to the high prevalence of CKD the actual number of poor outcomes attributable to CKD is higher. Since >50% of those over the age of 70 years have CKD, prediction models to stratify care by risk group, focusing on intervention, and delivering different models of care based on risk, have great potential particularly at a population level. Risk prediction models in CKD have been developed and show promise but, thus far, have limitations – clinical performance is not fully reported and external validation is rare. The clinical utility of these models lies, for example, in the ability to explore timing of dialysis access placement, but also requires further research. The introduction of such models has great potential to deliver appropriate stratified medical care, but this should be after appropriate randomised controlled trials of effect.

**REFERENCES**