




## CKJ REVIEW

## Translational research in nephrology: prognosis

Giovanni Tripepi<sup>1,\*</sup>, Davide Bolignano <sup>2,\*</sup>, Kitty J. Jager<sup>3</sup>, Friedo W. Dekker<sup>4</sup>, Vianda S. Stel<sup>3</sup> and Carmine Zoccali<sup>5,6,\*</sup>

<sup>1</sup>Institute of Clinical Physiology (IFC-CNR), Clinical Epidemiology and Physiopathology of Renal Diseases and Hypertension of Reggio Calabria, Italy, <sup>2</sup>Nephrology and Dialysis Unit, “Magna Graecia” University, Catanzaro, Italy, <sup>3</sup>Department of Medical Informatics, Academic Medical Center, Amsterdam Public Health Research Institute, University of Amsterdam, Amsterdam, The Netherlands, <sup>4</sup>Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands, <sup>5</sup>Renal Research Institute, New York, NY, USA and <sup>6</sup>Associazione Ipertensione, Nefrologia e Trapianto Renale (IPNET) c/o Nefrologia, Ospedali Riuniti, Reggio Calabria, Italy

\*These authors contributed equally to this work.

Correspondence to: Giovanni Tripepi; E-mail: [gtripepi@ifc.cnr.it](mailto:gtripepi@ifc.cnr.it)

### ABSTRACT

Translational research aims at reducing the gap between the results of studies focused on diagnosis, prognosis and therapy, and every day clinical practice. Prognosis is an essential component of clinical medicine. It aims at estimating the risk of adverse health outcomes in individuals, conditional to their clinical and non-clinical characteristics. There are three fundamental steps in prognostic research: development studies, in which the researcher identifies predictors, assigns the weights to each predictor, and assesses the model’s accuracy through calibration, discrimination and risk reclassification; validation studies, in which investigators test the model’s accuracy in an independent cohort of individuals; and impact studies, in which researchers evaluate whether the use of a prognostic model by clinicians improves their decision-making and patient outcome. This article aims at clarifying how to reduce the disconnection between the promises of prognostic research and the delivery of better individual health.

**Keywords:** calibration, discrimination, precision medicine, prognostic research

### INTRODUCTION

Medicine consists of diagnosis, prognosis, therapy and prevention. Thus, prognosis is a core element of clinical medicine. Prognosis aims at estimating the risk of health outcomes in various settings and diseases by investigating the association between risk factors (i.e. variables causally related to the occurrence of a given event) and biomarkers (i.e. predictors assessed at a given time point) with future clinical events in a certain population. The final aim of prognostic research is that

of identifying patients at relatively high risk of a given event by setting a proper scenario for improving health outcomes. Prognosticating outcomes is not synonymous with explaining their cause, this latter being a paradigm of exclusive pertinence to aetiological research.

As the world population grows, so does the number of individuals affected by disease, which makes prediction of future events fundamental for tailoring the clinical surveillance in a specific patient category in order to obtain better outcomes. Prognostic research is integral to the clinical decision process

Received: 30.6.2021; Editorial decision: 10.8.2021

© The Author(s) 2021. Published by Oxford University Press on behalf of the ERA-EDTA.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

in healthcare. However, a large gap still exists between the potential of prognostic research and the actual impact that this research makes in the real world of clinical medicine.

Prognostic research demands: (i) the definition of populations, expected health outcomes for a given condition in a given country or world area in a certain period; (ii) the identification of candidate risk factors and biomarkers; (iii) the development and validation of risk prediction models; and (iv) the application of prognostic information for treatment decisions at individual level (personalized care) or for clusters of patients characterized by shared clinical characteristics or risk factors (stratification) (Figure 1).

## PROGNOSIS RESEARCH: THE BASICS

The scope of predictive modelling is to identify the likelihood of future events, and such a modelling is typically used in meteorology to predict the weather. This is a multivariate procedure based on a set of climate variables like barometric pressure, clouds, winds, temperature, time of the year and other variables. From these variables, weather forecasters make reasonably accurate predictions at local level, i.e. in a given city or in a given region. By the same token, predictive models in medicine are developed to predict clinical outcomes [e.g. death, cardiovascular (CV) or renal events] in patients with a given disease. Predictive models are also used in screening programmes to identify individuals at high risk for various conditions such as cardiac ischaemia or neoplasia or chronic kidney disease (CKD).

Causal modelling is a cornerstone in epidemiology for identifying causal risk factors. In contrast, predictive modelling just aims at predicting future events by considering variables independently of their link (causal or non-causal) with outcomes. Prediction and explanation are often erroneously conflated concepts because both encompass the term 'risk factor'. Regression analysis, e.g. Cox's regression, is applied both for causal inference and prediction. However, there are some important differences in how these techniques are used in these applications. For example, in causal research omission of confounding variables may produce biased estimates of the regression coefficients and therefore invalidate the models. Known variables that both affect the outcome variable and the risk factor to be studied (confounders) are crucial to adjust for in causal modelling. Omission of such variables can lead to erroneous causal conclusions. In contrast, for predictive modelling omitting one or more variables is not a problem as long as these variables are not critical to optimize predictions. In other words, predictive models optimize predictions by combining available variables. In this respect, as long as they may serve to improve data fitting, interactions can be included to improve the models. However, these

interactions *per se* have no meaning or implication. Another difference between causal and predictive models is the fact that the explained variance of the outcome variable ( $R^2$ ) is of utmost importance in predictive studies but not for studies on causal inference. In simple terms, in these models, the standard error of the prediction, a metric related to the  $R^2$ , is crucial in predictive research. The higher the  $R^2$ , the better the predictive model. In contrast, the  $R^2$  may not be important in causal research. Indeed, a low  $R^2$  may be compensated by a large sample size.

As remarked, risk prediction models do not necessarily include causal risk factors and may include costless and easy to capture variables like the condition of widower in a survival model in the elderly. In general, these models are based on statistical criteria that maximize the fit of the data to the outcome. This can be done by stepwise selection of variables, and increasingly so in recent times, by machine learning methods. Several 'risk factors' are used to obtain a model that efficiently fits the data and that explains the greatest possible proportion of variance in the outcome of interest. Once developed, prognostic models need to be externally validated in populations similar to the one where they were developed. Indeed, these models are typically population-specific and almost always they do not perform well in other populations [1-2].

## PROGNOSTIC FACTORS RESEARCH

Prognostic factors, by definition, are measures able to predict the occurrence of particular outcomes in individuals. Blood pressure and albuminuria predict CV endpoints in the general population and even more in CKD individuals [3]; CKD itself [reduced estimated glomerular filtration rate (eGFR)] is a prognostic biomarker for CV events in virtually all clinical settings [4]. However, with the notable exception of age, single risk factors on their own explain just a limited proportion of the variance of future clinical events. This is true for all risk factors mentioned above, as well as for serum cholesterol and diabetes. Only if properly combined into multivariate models do these risk factors efficiently predict clinical outcomes. Framingham risk factors, when combined, explain over 70% of the variance of incident coronary heart disease but are *per se* weak predictors when used individually [5].

Prognosis might be used to guide prevention. For example, since the relationship between low density lipoprotein (LDL) cholesterol and the risk of cardiac ischaemia is linear, the clinical decision to reduce lipids is not based on the diagnosis of hyperlipidaemia (LDL cholesterol >160 mg/dL) but rather on the absolute risk of the condition. Thus, in individuals who have suffered from myocardial infarction (MI) and who are at high risk for a second episode of the same event, a lipid-lowering drug is recommended, independently of the presence of hyperlipidaemia, to reach LDL levels of <70 mg/dL, which is a very low level of LDL cholesterol, about the 5th percentile of the distribution of this biomarker in the general population. However, we should be clear that strictly speaking the knowledge we have that taking statins lowers the actual risk of a second MI originates from well-designed aetiological studies, and especially from successful randomized controlled trials, and not from the fact that cholesterol levels were included in a prediction models for MI.

Beyond the application of risk prediction at individual level, prognostic research allows the comparison of prognosis of diseases across countries, and such a comparison is useful to have an insight into the effectiveness of cares by various health systems on the same diseases. In 2014, Sheng-Chia Chung et al. [6]

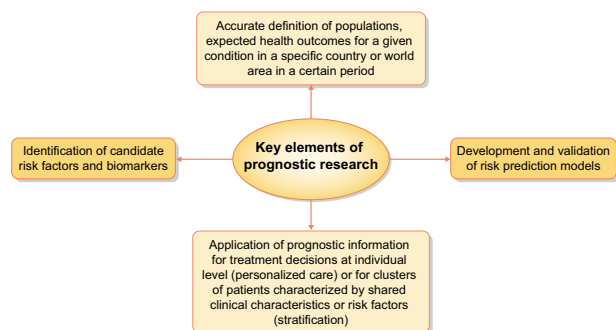


FIGURE 1: Key elements to adequately address a prognostic research.

compared mortality at 30 days post-MI in 119786 patients in Sweden and 391077 in the UK. These two countries were specifically selected because they have similar health systems and invest similar financial resources for acute MI. This comparison showed much higher unadjusted mortality in the UK [10.5%; 95% confidence interval (CI) 10.4–10.6] than in Sweden (7.6%; 7.4–7.7), and the UK to Sweden standardized mortality ratio was 1.37 (1.30–1.45). This important inter-country difference in outcomes post-MI was attributed by these authors to the fact that diffusion of evidence-based changes to practice and new technologies is quicker in Sweden. Thus, variables like health system organization, identified by the corresponding country, which are not prognostic on their own when applied at country level may become quite strong predictors in multivariate models including patients with the same condition across countries.

Multiple prognostic factors may be combined in a prognostic model with in an attempt to optimize individual risk prediction [7] (see next section). For instance, the Framingham risk score in the USA (developed in Framingham study participants who attended a routine examination between 30 and 74 years of age and were free from CV diseases) [8], the Systematic COronary Risk Evaluation (SCORE) in Europe [developed in people without overt CV disease, diabetes (Types 1 and 2), CKD, familial hypercholesterolaemia or very high levels of individual risk factors] [9] and the Global Registry of Acute Coronary Events (GRACE) [10] risk score (developed in patients hospitalized with acute coronary syndromes) are widely adopted for predicting CV outcomes in the general population and in hospitalized coronary patients, respectively. A renal risk score (based on eGFR, age, urinary albumin excretion, systolic blood pressure (BP), C-reactive protein and known hypertension) was developed to identify individuals at increased risk for developing progressive CKD [11]. The Malnutrition-Inflammation Score (MIS), which combines information from the nutritional status and the inflammatory response, predicts death and poor outcomes in CKD and in dialysis patients [12]. In almost all cases, prognostic factors are biomarkers, i.e. biological, anthropometric, imaging or physiological variables that can be measured in routine clinical practice. Nephrologists are familiar with serum creatinine, proteinuria and/or albuminuria. Other biomarkers like body mass index and waist-to-hip ratio in obesity or NMR estimates of renal volume in autosomal dominant polycystic kidney disease (ADPKD) also qualify as prognostic factors because they are useful for risk stratification in these conditions. On the other hand, social status, income, healthcare access, environmental factors and geography are predictive factors of paramount relevance in that these factors explain the variability in health outcomes even more than physiological or pathophysiological parameters. Lower socio-economic status predicts incident CKD and progression of CKD to end-stage kidney disease and poor health outcomes [13]; the geographical area of origin predicts access to dialysis resources and transplantation [14]. Needless to say, socio-economic variables have *per se* no direct causal link with health conditions. However, they capture the risk associated with low economic status, like poor nutrition and hygienic conditions, and exposure to unhealthy environments.

In certain pathological conditions, there might be poor or insufficient knowledge on the biological mechanisms underlying the evolution of the disease and, consequently, few or no clues for deciding the set of biomarkers to include in predictive models. Studies dealing with such conditions usually blindly screen a wide series of factors (e.g. genetic polymorphisms or psychosocial factors), some of them unsuspected or without biological plausibility, and test their possible association with the

outcome. This so-called ‘biology agnostic’ approach is being extensively adopted in prognostic research in nephrology, mostly after the advent of ‘omic’ techniques, such as genomics (DNA), transcriptomics (RNA), proteomics (proteins) or metabolomics (metabolites).

Once potential prognostic factors are identified in one study, early replication in multiple independent studies and comparisons with existing predictive models are crucial steps towards the applicability in clinical routine. In fact, a biomarker may perform as a single excellent prognostic factor but its addition to other existing predictive models may not improve the overall prognostic information. For instance, Urinary Neutrophil Gelatinase-Associated Lipocalin (uNGAL) has shown an excellent ability for the early detection and the prognosis of acute kidney injury [15]. Yet, in a study conducted on a heterogeneous population of 451 critically ill adults [16], the area under the curve (AUC) of uNGAL for acute kidney injury detection was lower than that of a simple prognostic model including the APACHE-II score plus serum creatinine, presence of sepsis and clinical department of origin (AUC: 0.71 versus 0.81). Furthermore, the addition of uNGAL did not improve the overall prognostic accuracy of such a model (AUC: 0.81 in the model without uNGAL and 0.82 in the model including uNGAL), indicating that the measurement of this biomarker, which is not a cheap biomarker, may be omitted with minimal or no loss of information for the risk assessment in the intensive care unit setting.

As briefly alluded to before, the eventual aim of prognostic research is to change (improve) clinical practice and patient management. In modern medicine, diagnostic criteria and staging of diseases are continuously refined. This is mostly consequent to the advent of new biomarkers or to the elaboration of new prognostic models. Old definitions of CKD relied on the presence of an overt impairment in renal function (decrease in eGFR). Prognostic studies have subsequently elucidated that the sole presence of stable haematuria or persistently abnormal protein urine excretion indicates individuals at high risk of adverse renal and CV outcomes in the long-term, even in the presence of eGFR within the normal range. Current diagnostic criteria and CKD staging definitions (KDOQI) have implemented this prognostic information and subjects with persistent urinary abnormalities (but normal eGFR) are now labelled as having ‘kidney injury’ (CKD Stage 1) [17]. One important clinical applicability of prognostic variables (i.e. of variables that are not necessarily causally implicated in the pathway of the disease) is to monitor change in disease status and the response to therapy overtime. Reduction or normalization of proteinuria in subjects with membranous nephropathy suggests successful treatment and improvement in renal damage [18]. By the same token, albuminuria reduction in hypertensive subjects by proper treatment (e.g. renin-angiotensin system blockers) decreases their overall CV risk [19].

Prognostic factors may also drive the development of new interventions, or new applications of existing interventions, when the relationship between the risk factor and the following outcome is causal in nature. Not rarely, causal risk factors are only weakly related to adverse health outcomes, as it is the case of serum cholesterol for MI. In most cases, prognostic factors are mere risk markers, i.e. are factors not directly involved in the pathogenesis of diseases. For example, tumour markers predict cancer recurrence but most of them are simply epiphenomena of the disease.

The number of prognostic studies is growing steadily but the quality of most studies remains suboptimal. Quite surprisingly,

no standardized methods on how to conduct prognostic research have been definitely established. Yet, prognostic studies are often poorly designed, inappropriately analysed or barely reported. Furthermore, replication of initial findings is often inadequate, and the overall evidence may be tainted by a high risk of publication bias and selective reporting. For these reasons, confusion may arise about the true prognostic value of risk factors being tested. In 2013, the Prognosis Research Strategy (PROGRESS) group was formed to improve the quality of prognostic research by establishing standards and providing recommendations on how such research should be conducted [20]. The group stressed the necessity of large, prospective, registered and protocol-supported studies with adequate sample size, proper statistical analyses and transparent reporting of all factors and outcomes considered. According to the recommendations by PROGRESS, exploratory studies should be clearly labelled as such and the global prognostic ability of a given factor should be examined by pooling data of multiple studies in meta-analysis to alleviate any possible reporting bias and information deficiency in the original studies. Finally, for each factor recognized as prognostic, there should be clear awareness of how this can be used for improving clinical outcomes, including whether it might be valuable for the clinical management of patients or might rather provide new insights for interventional research. The Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement for the reporting of prediction models provides precious recommendations for guiding research looking at predictive models for prognosis or diagnosis [21].

## PROGNOSTIC MODELS RESEARCH

The development and the validation of prognostic models represent an important research area of clinical epidemiology, and well-validated risk prediction rules are fundamental tools for decision-making in clinical medicine. The development of a risk prediction model for a given event [death, MI, stroke, end-stage kidney failure (ESKF), etc.] demands the identification of prognostic factors (i.e. variables that predict the event of interest

with high accuracy) in well-characterized cohorts of subjects/patients followed up for a given period of time. Once identified in a given cohort, the prognostic factors are combined into a mathematical equation by estimating for each variable a regression coefficient that maximizes the prediction accuracy of the event of interest [22]. The accuracy of a risk prediction model for predicting an event is assessed by measuring calibration, discrimination and risk reclassification.

### Calibration and discrimination

Calibration measures how much the prognostic estimate of a specific predictive model including one or more clinical characteristics/prognostic markers matches the 'real' probability of the outcome (i.e. the observed proportion of an event in a given period of time) [23], whereas discrimination expresses the ability of the risk prediction model to distinguish individuals with from those without the event of interest [24] (Figure 2). The discriminatory power of prognostic models fitted to survival data (i.e. data including censored observations) is commonly assessed by calculating the Harrell's C statistic [24], which is conceptually similar to a receiver operating characteristic curve analysis [25]. The higher the C-index, the higher the prognostic accuracy of the model. Discrimination, as assessed by the Harrell's C-index, ranges from 0.5 (no discrimination) to 1.0 (perfect discrimination).

### Akaike information criterion, likelihood ratio test and risk reclassification analysis

Another index is the Akaike information criterion (AIC) [26], a method that is used to compare non-nested prognostic models (Figure 2). Two models are nested if one model can be reduced to the other one by imposing a restriction in the number of predictors being tested. For example, given a model (e.g. a Model X) including three prognostic variables (A, B and C), a nested model is that including 'A and B' or 'B and C' or 'A and C'. Vice versa, two models are non-nested if one model cannot be reduced to the other model by imposing a restriction in the number of predictors being tested (e.g. a non-nested model of Model X

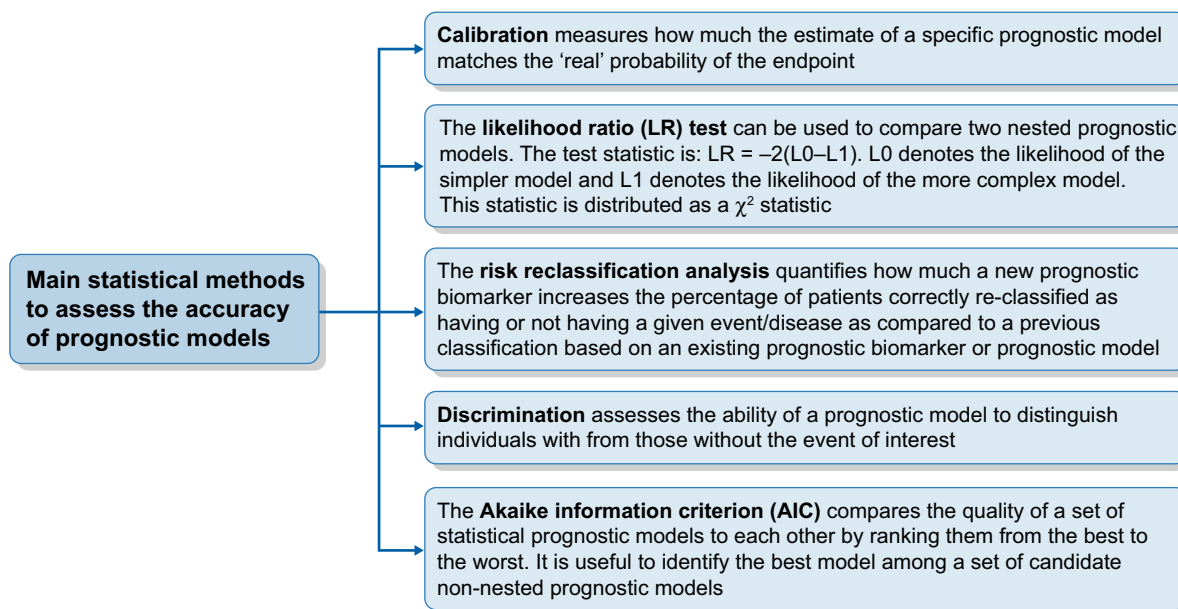


FIGURE 2: Main statistical methods to assess the accuracy of prognostic models.

mentioned above is that including as predictors 'A, B and D'. Given two non-nested models, the preferred model to predict survival (i.e. that having the best performance) is the one with the lowest AIC value. The advantage of AIC is that it includes a penalty that is an increasing function of the number of predictors. The penalty clearly discourages over-fitting in order to develop parsimonious models. To compare two nested prognostic models, the likelihood ratio (LR) test can be used (Figure 2). The test statistics is  $LR = -2(L_0 - L_1)$ .  $L_0$  denotes the likelihood of the simpler model and  $L_1$  denotes the likelihood of the more complex model. This statistics is distributed as a chi-square statistics. An important step in assessing the performance of a prognostic model is the analysis of risk reclassification, which is assessed by calculating the Net Reclassification Index (NRI) [23] (Figure 2). Such an index provides a direct measure of the impact of new biomarkers in risk predictions. Reclassification quantifies how much a new prognostic biomarker increases the percentage of patients correctly reclassified as having or not a given event/disease as compared with a previous classification based on an existing prognostic biomarker or predictive model. No specific reference values are available for the AIC and the NRI. The statistical techniques used to develop a prognostic model (discrimination, calibration and risk reclassification) are exactly the same as those used to validate a risk prediction equation. A model can be internally and externally validated. The most common method for internal validation is the cross-validation. By this method, the original cohort is split into a development and a validation sample. The model-building procedures applied to the development sample (i.e. discrimination, calibration and risk reclassification analyses) are tested in the validation sample for assessing reproducibility. The external validation examines the generalizability of a risk prediction model to completely independent cohorts by using the full set of state-of-art statistical methods such as discrimination, calibration and risk reclassification.

## PROGNOSTIC MODELS IN CLINICAL PRACTICE

CKD is a progressive condition and early identification of patients at high risk of developing ESKF is fundamental to slowing disease progression, maintaining quality of life and improving outcomes. Tangri *et al.* [27] developed a risk equation [the 5-year Kidney Failure Risk Equation (KFRE)] for predicting progression of CKD to ESKF (dialysis/kidney transplantation) in a very large series of CKD Stages 3–5 patients. They used data from two independent CKD cohorts: a development cohort including 3449 CKD patients and a validation cohort including 4942 CKD patients. Patients in the two cohorts were quite similar for demographic, clinical and biochemical data. In the development cohort, as candidate prognostic factors the authors considered baseline eGFR as well as age, gender, BP, weight, diabetes, hypertension and CKD aetiology, and laboratory data such as urine albumin to creatinine ratio (ACR), serum albumin, serum phosphate, serum bicarbonate and serum calcium. All variables that resulted to be associated with ESKF at univariate Cox regression analyses (with  $P < 0.10$ ) were combined in risk equations of various complexity in order to identify the best prognostic model with the minimum set of predictors. To do this, Tangri *et al.* calculated, for each candidate model, two indexes of prognostic performance, namely the C-index and the AIC. A predictive model including age and gender (Basic model) provided a poor discriminatory power for ESKF (only 56%), indicating that these two variables were inaccurate to estimate renal prognosis in CKD patients. The inclusion of baseline eGFR

and albuminuria (Model 1) increased the accuracy in prediction of the basic model (C-index from 89% to 91%,  $P < 0.001$ ; AIC from 4834 to 4520), indicating that these two risk factors play an important role for predicting the evolution of CKD toward ESKF. By adding into Model 1, serum albumin, serum phosphate, serum bicarbonate and serum calcium, both the C-index (from 91% to 92%,  $P < 0.001$ ) and the AIC (from 4520 to 4432) slightly improved. Forcing into this model the remaining prognostic factors that resulted to be associated with ESKF at univariate Cox regression analyses (diabetes, hypertension, BP and body weight) did not improve the discriminatory power of the risk equation, indicating that a prediction model including eight variables (namely age, sex, eGFR, albuminuria, calcium, phosphate, bicarbonate and albumin) is the best one, among the set of candidate models, for predicting the risk of ESKF in CKD patients. By combining these risk factors, the authors built up a risk calculator (in Excel format) that is easily downloadable from the web (<http://jama.jamanetwork.com/article.aspx?articleid=897102>). To explain how this risk calculator can be used, we consider a 50-year-old man with CKD (eGFR of 30 mL/min/1.73 m<sup>2</sup>), an ACR of 50 mg/g, a serum calcium of 9.8 mg/dL, a serum phosphate of 3.8 mg/dL, a serum albumin of 4 g/dL and a serum bicarbonate of 26 mEq/L. In this patient, the 5-year probability of ESKF estimated by the software is 11%. If we want to assess how much the risk of ESKF could increase in the same man if the level of serum phosphate were to increase from 3.8 mg/dL to 6.0 mg/dL, we might calculate the risk of ESKF associated with a value of serum phosphate of 6.0 mg/dL (instead of 3.8 mg/dL) by leaving unchanged the remaining prognostic factors. By doing this, we find an estimated 5-year risk of ESKF of 17%, a figure substantially higher than that found for the same man with normal serum phosphate (11%). The eight-variables risk equation was externally validated by the authors in an independent cohort of 4942 CKD patients. The discriminatory power of the prognostic model in the validation cohort was very satisfactory (C-index: 84%, 95% CI 83–86%) indicating that the risk calculator has external generalizability for predicting the risk of ESKF in CKD patients. In the validation cohort, the risk reclassification analysis was stratified by CKD stages and performed by comparing the eight-variables prognostic model (expanded model) with a reduced model including four variables (i.e. age, gender, ACR and eGFR). This analysis showed that the expanded model was more accurate than the reduced one for reclassifying patients in all CKD stages, and this was particularly true in CKD Stage 3, where the NRI was 8.0% (95% CI 2.1–13.9%). An 8% NRI indicates that the use of the expanded model produced an 8% net improvement in reclassification as compared with the use of the reduced model ( $P < 0.05$ ). In more detail, the expanded model as compared with the reduced model provided an 8.5% improvement in risk classification in patients with ESKF and a 0.5% worsening in risk classification in patients without ESKF, thus generating an 8% net improvement in risk accuracy estimation. The authors concluded that the eight-variable prognostic model accurately predicts progression to kidney failure in patients with CKD Stages 3–5.

The Tangri model apart, several kidney failure prediction models have been published over the last 10 years. However, these models were developed and validated in cohorts of patients with a wide range of disease severity, without accounting for the competing risk of death [28]. Furthermore, currently used prediction models have not been compared for accuracy and precision. In a freshly published study [28] aimed at externally validating 11 existing models of kidney failure, most models with longer prediction horizons largely overestimated the

risk of kidney failure. The 5-year KFRE [27], which does not consider the competing death risk, overpredicted by 10–18% the risk of kidney failure, whereas the 2-year KFRE [2] (another equation by Tangri specifically developed to predict the risk of kidney failure in a relatively short time period and that does not account for the competing mortality risk) performed well over a 2-year time frame, indicating that taking the competing death risk into account is not necessary for short-term risk predictions in patients with advanced CKD. The 4-year Grams [29] model, which includes nine variables (age, eGFR, sex, race, CV disease, diabetes, systolic BP, uACR and smoking) and accounts for the competing death risk, showed excellent calibration and good discrimination for predicting kidney failure in patients with CKD and severely decreased GFR. Thus, taking into consideration the competing risk of mortality is critical for long-term but not for short-term predictions.

## PROGNOSTIC RESEARCH AND PRECISION MEDICINE

Precision medicine (also known as personalized medicine) is a new approach based on dividing patients into groups (stratification) according to their biological characteristics, the risk of developing diseases and, mostly, the chance of being respondent to particular therapies [30]. The ultimate goal of precision medicine is to individualize therapy and make the best decisions for comparable groups of patients, providing the 'right treatment, for the right person, at the right time'. Hence, this approach is posed in contrast to empirical medicine. Precision medicine is now acknowledged as a key global priority for pharmacologic and diagnostic industries and, above all, for healthcare providers [31]. Prognostic research is the backbone of precision medicine. Prognostic models may allow risk stratification and help in identifying priority areas for research. For example, they might clarify whether a given intervention associates with evident risks or greater costs, or whether inter-individual variations in the pharmacokinetics and metabolism of a drug may lead to different pharmacological responses. No less important, prognostic research may elucidate the presence of clinically important differences in the individual prognosis. For example, renal artery stenosis is commonly found among elderly patients with generalized atherosclerosis. Hypertension, which is also highly prevalent in these subjects, does not call for renal artery revascularization as most cases are essential in nature and the individual prognosis and response to anti-hypertensive agents are excellent. Conversely, sudden hypertension in middle-aged women with renal stenosis due to fibromuscular dysplasia is a compelling indication for revascularization as these subjects are more likely to develop refractory hypertension, and poor CV and renal outcomes [32]. Prognostic factors may help in personalizing treatments by estimating the absolute risk for one or more outcomes. Indeed, individuals with the highest absolute risk would take the greatest benefit from an intervention, as they would experience the greatest reduction in probability of the outcome. Such intervention might therefore be restricted only to particular strata of patients, to minimize adverse events and to optimize cost-effectiveness and chance of success. All major nephrology guidelines advocate the achievement and maintenance of haemoglobin levels within an optimal range, using erythropoiesis-stimulating agents (ESAs) if needed, as this may reduce adverse CV outcomes [33, 34]. By the same token, biomarker evidence of scarce iron reserve or high inflammatory activity predicts poor response to ESAs, demanding

appropriate correction therapy before considering administration of these drugs [35]. Precision approaches may also be useful when the treatment effect differs among patients groups. Differences in the response to treatment might depend on an interaction between the variable and the effect of treatment on the outcome. In this regard, when a plausible biological mechanism exists a precision approach may be applied to screen individuals for the presence of factors that may forecast successful response to therapies or interventions, in terms of more benefit, less harm or both. For example, normo- or micro-albuminuric diabetics carrying the II genotype of the *angiotensin-converting enzyme* (ACE) gene are particularly responsive to ACE inhibitors in terms of delayed progression of nephropathy, whereas the DD genotype is associated with a better response to ARBs therapy in Type II diabetics with overt nephropathy and to ACE inhibitors therapy in male patients with non-diabetic proteinuric nephropathies [36]. Although precision medicine relies on individual baseline information for tailoring therapeutic decisions, this approach is conceptually different from stepped or adaptive models of clinical care [37]. In these models, therapeutic adaptations (which can be represented by changes in dose scheme, duration or regimen) are influenced by the response or non-response to treatments previously administered. An example of such approach is represented by treatment of idiopathic nephrotic syndrome. Patients suffering from this condition are first prescribed a therapeutic cycle of corticosteroids to which the majority is deemed to be respondent. Hence, more aggressive solutions (e.g. immunosuppressive treatments) are reserved only for those non-respondent to this first attempt [38].

It is important to emphasize that not all prognostic factors are also predictors of treatment response, and vice versa. For example, response to statin treatment has been associated with polymorphisms in several genes such as the cholesterol ester transfer protein, 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase, apolipoprotein E and apolipoprotein-CI. However, although these genes modulate the entity of decrease in lipid levels during statin therapy, *per se* these genes are unrelated to CV risk [39]. Several challenges and research limitations currently hinder a wide applicability of precision approaches, particularly in certain fields of medicine such as nephrology. False-negative findings (also known as Type II errors) may lead to inappropriate conclusions such as considering a particular factor as not prognostically useful when actually it is [40]. For instance, poorly designed randomized trials may have inadequate statistical power for detecting factors that are truly prognostic for different treatment effects. Therefore, the aim of prognostic analyses should be clearly explained in these studies and appropriately considered for the calculation of study power. Similarly, dichotomization of continuous variables or arbitrary data dredging may reduce power further [41]. Indeed, factors have more power and less risk of bias when analysed on a continuous scale than after dichotomization according to a cut-off. Firm prognostic evidence may thus appear gradually from secondary analyses of such trials or from meta-analyses pooling individual participant data [42]. Likewise, false-positive findings (Type I errors) may come out from inappropriate use of subgroup analyses. False positives may arise by chance if proper multiple statistical analyses are not implemented, particularly in the presence of a large number of factors to consider [43]. Furthermore, the selection of endpoints itself can lead to misinterpretation of the prognostic value of a given factor in predicting response to treatment. Indeed, positive findings are more likely reported when looking at surrogate endpoints of treatment effects, rather than hard, patient-centred outcomes such

as mortality, major events or disease progression [44]. In pilot studies, as well as in confirmation trials, mTOR inhibitors showed good capacity of reducing total kidney and cyst volumes in ADPKD patients (a surrogate endpoint of disease severity) but failed to halt disease progression in terms of eGFR decline over time [45]. In conclusion, if a treatment effect is deemed inconsistent across individuals, a personalized approach based on rigorous tests for predicting treatment response may guide treatment decisions. Unfortunately, the application of precision medicine in nephrology is currently marginal, as strong evidence of clinical impact is limited and spurious findings on predictive factors prevail over externally well-conducted and validated prognostic studies. Improvement in the way prognostic research is conducted, including cost-evaluation analyses of implementing precision approaches, is mandatory to fill the existing gap between nephrology and other medical specialties.

## CONCLUSIONS

Knowledge of the clinical outcomes of diseases is fundamental in order to optimize the clinical decision-making process. The relevance of such knowledge impacts upon various translational pathways. Prognostic research is of paramount importance for precision medicine, a new approach that aims to individualize therapy in order to provide 'the right treatment, for the right person, at the right time'. Risk calculators allow informing patients of the future course of their disease and help clinicians in formulating a prognosis. In order to judge applicability of risk prediction models in clinical practice, risk calculators should be appropriately validated by providing measures of calibration, discrimination and risk reclassification. The definitive proof of the clinical usefulness of a risk prediction model rests on the demonstration that a treatment policy guided by the risk calculator is more effective than that provided by the standard care for reducing the progression rate of CKD, an issue which should be specifically tested in a randomized clinical trial.

## CONFLICT OF INTEREST STATEMENT

None declared. The results presented in this article have not been published previously in whole or part, except in abstract format.

## REFERENCES

- Pennells L, Kaptoge S, Wood A *et al.*; Emerging Risk Factors Collaboration. Equalization of four cardiovascular risk algorithms after systematic recalibration: individual-participant meta-analysis of 86 prospective studies. *Eur Heart J* 2019; 40: 621–631
- Tangri N, Grams ME, Levey AS *et al.*; CKD Prognosis Consortium. Multinational assessment of accuracy of equations for predicting risk of kidney failure. A meta-analysis. *JAMA* 2016; 315: 164–174
- Weir MR. Microalbuminuria and cardiovascular disease. *Clin J Am Soc Nephrol* 2007; 2: 581–590
- Sarnak MJ, Levey AS, Schoolwerth AC *et al.*; Clinical Cardiology, and Epidemiology and Prevention. Kidney disease as a risk factor for development of cardiovascular disease: a statement from the American Heart Association Councils on Kidney in Cardiovascular Disease, High Blood Pressure Research. *Hypertension* 2003; 42: 1050–1065
- Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007; 115: 928–935
- Chung SC, Gedeberg R, Nicholas O *et al.* Acute myocardial infarction: a comparison of short-term survival in national outcome registries in Sweden and the UK. *Lancet* 2014; 383: 1305–1312
- Bolignano D, Mattace-Raso F, Torino C *et al.* Prognostic models in the clinical arena. *Aging Clin Exp Res* 2012; 24: 300–304
- D'Agostino RB, Vasan RS, Pencina MJ *et al.* General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* 2008; 117: 743–753
- 2019 ESC/EAS Guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk. *Eur Heart J* 2020; 41: 111–188
- GRACE Investigators. Rationale and design of the GRACE (Global Registry of Acute Coronary Events) Project: a multinational registry of patients hospitalized with acute coronary syndromes. *Am Heart J* 2001; 141: 190–199
- Halbesma N, Jansen DF, Heymans MW *et al.*; for the PREVEND Study Group. Development and validation of a general population renal risk score. *Clin J Am Soc Nephrol* 2011; 6: 1731–1738
- Rambod M, Kovesdy CP, Kalantar-Zadeh K. Malnutrition-Inflammation Score for risk stratification of patients with CKD: is it the promised gold standard? *Nat Clin Pract Nephrol* 2008; 4: 354–355
- McClellan WM, Newsome BB, McClure LA *et al.* Poverty and racial disparities in kidney disease: the REGARDS study. *Am J Nephrol* 2010; 32: 38–46
- Crews DC, Liu Y, Boulware LE. Disparities in the burden, outcomes, and care of chronic kidney disease. *Curr Opin Nephrol Hypertens* 2014; 23: 298–305
- Haase M, Bellomo R, Devarajan P *et al.*; NGAL Meta-analysis Investigator Group. Accuracy of neutrophil gelatinase-associated lipocalin (NGAL) in diagnosis and prognosis in acute kidney injury: a systematic review and meta-analysis. *Am J Kidney Dis* 2009; 54: 1012–1024
- Siew ED, Ware LB, Gebretsadik T *et al.* Urine neutrophil gelatinase-associated lipocalin moderately predicts acute kidney injury in critically ill adults. *J Am Soc Nephrol* 2009; 20: 1823–1832
- KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. *Kidney Int Suppl* 2013; 3: 1
- Glasscock RJ. The pathogenesis of membranous nephropathy: evolution and revolution. *Curr Opin Nephrol Hypertens* 2012; 21: 235–242
- Pascual JM, Rodilla E, Costa JA *et al.* Prognostic value of microalbuminuria during antihypertensive treatment in essential hypertension. *Hypertension* 2014; 64: 1228–1234
- Hemingway H, Croft P, Perel P *et al.*; PROGRESS Group. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ* 2013; 346: e5595
- Moons KG, Altman DG, Reitsma JB *et al.* Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015; 162: W1–W73
- Tripepi G, Heinze G, Jager KJ *et al.* Risk prediction models. *Nephrol Dial Transplant* 2013; 28: 1975–1980
- Tripepi G, Jager KJ, Dekker FW *et al.* Statistical methods for the assessment of prognostic biomarkers(part II): calibration and re-classification. *Nephrol Dial Transplant* 2010; 25: 1402–1405

24. Tripepi G, Jager KJ, Dekker FW et al. Statistical methods for the assessment of prognostic biomarkers (Part I): discrimination. *Nephrol Dial Transplant* 2010; 25: 1399–1401
25. Tripepi G, Jager KJ, Dekker FW, Zoccali C. Diagnostic methods 2: receiver operating characteristic (ROC) curves. *Kidney Int* 2009; 76: 252–256
26. Steyerberg EW. *Clinical Prediction Models. A Practical Approach to Development, Validation, and Updating*. New York: Springer, 2009
27. Tangri N, Stevens LA, Griffith J et al. A predictive model for progression of chronic kidney disease to kidney failure. *JAMA* 2011; 305: 1553–1559
28. Ramspek CL, Evans M, Wanner C et al.; EQUAL Study Investigators. Kidney failure prediction models: a comprehensive external validation study in patients with advanced CKD. *J Am Soc Nephrol* 2021; 32: 1174–1186
29. Grams ME, Sang Y, Ballew SH et al. Predicting timing of clinical outcomes in patients with chronic kidney disease and severely decreased glomerular filtration rate. *Kidney Int* 2018; 93: 1442–1451
30. Trusheim MR, Berndt ER, Douglas FL. Stratified medicine: strategic and economic implications of combining drugs and clinical biomarkers. *Nat Rev Drug Discov* 2007; 6: 287–293
31. Hamburg MA, Collins FS. The path to personalized medicine. *N Engl J Med* 2010; 363: 301–304
32. Safian RD. Atherosclerotic renal artery stenosis. *Curr Treat Options Cardiovasc Med* 2003; 5: 91–101
33. Horl WH. Anaemia management and mortality risk in chronic kidney disease. *Nat Rev Nephrol* 2013; 9: 291–301
34. Drueke TB, Parfrey PS. Summary of the KDIGO guideline on anemia and comment: reading between the (guide)line(s). *Kidney Int* 2012; 82: 952–960
35. Tarnag DC, Huang TP, Chen TW et al. Erythropoietin hyporesponsiveness: from iron deficiency to iron overload. *Kidney Int* 1999; 55: 107–118
36. Ruggenti P, Bettinaglio P, Pinares F et al. Angiotensin converting enzyme insertion/deletion polymorphism and renoprotection in diabetic and nondiabetic nephropathies. *Clin J Am Soc Nephrol* 2008; 3: 1511–1525
37. Almirall D, Compton SN, Gunlicks-Stoessel M et al. Designing a pilot sequential multiple assignment randomized trial for developing an adaptive treatment strategy. *Stat Med* 2012; 31: 1887–1902
38. Waldman M, Austin HA. Treatment of idiopathic membranous nephropathy. *J Am Soc Nephrol* 2012; 23: 1617–1630
39. Yip VL, Pirmohamed M. Expanding role of pharmacogenomics in the management of cardiovascular disorders. *Am J Cardiovasc Drugs* 2013; 13: 151–162
40. Yusuf S, Wittes J, Probstfield J et al. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA* 1991; 266: 93–98
41. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med* 2006; 25: 127–141
42. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010; 340: c221
43. Assmann SF, Pocock SJ, Enos LE et al. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet* 2000; 355: 1064–1069
44. Holmes MV, Shah T, Vickery C et al. Fulfilling the promise of personalized medicine? Systematic review and field synopsis of pharmacogenetic studies. *PLoS One* 2009; 4: e7960
45. He Q, Lin C, Ji S et al. Efficacy and safety of mTOR inhibitor therapy in patients with early-stage autosomal dominant polycystic kidney disease: a meta-analysis of randomized controlled trials. *Am J Med Sci* 2012; 344: 491–497