THE NEED FOR ACCURATE RISK PREDICTION MODELS FOR ROAD MAPPING, SHARED DECISION MAKING AND CARE PLANNING FOR THE ELDERLY WITH ADVANCED CHRONIC KIDNEY DISEASE

Marijke Stryckers, Evi V Nagler, Wim Van Biesen

Renal Division Ghent University Hospital, Ghent Belgium

Corresponding Author: Evi V.T. Nagler, Renal Division, Department of Internal Medicine, Ghent University Hospital, De Pintelaan 185, 9000 Gent, Belgium; tel: +32 9 332 0417; fax: +32 9 332 3847; e-mail: evi.nagler@ugent.be

Abstract
As people age, chronic kidney disease becomes more common, but it rarely leads to end-stage kidney disease. When it does, the choice between dialysis and conservative care can be daunting, as much depends on life expectancy and personal expectations of medical care. Shared decision making implies adequately informing patients about their options, and facilitating deliberation of the available information, such that decisions are tailored to the individual’s values and preferences. Accurate estimations of one’s risk of progression to end-stage kidney disease and death with or without dialysis are essential for shared decision making to be effective. Formal risk prediction models can help, provided they are externally validated, well-calibrated and discriminative; include unambiguous and measureable variables; and come with readily applicable equations or scores. Reliable, externally validated risk prediction models for progression of chronic kidney disease to end-stage kidney disease or mortality in frail elderly with or without chronic kidney disease are scant. Within this paper, we discuss a number of promising models, highlighting both the strengths and limitations physicians should understand for using them judiciously, and emphasize the need for external validation over new development for further advancing the field.

Keywords: Prognosis; Proportional Hazard models; Logistic Models; Aged; Renal Insufficiency, Chronic

Introduction
The prevalence of Chronic Kidney Disease (CKD) increases with age, but few of the elderly actually progress to End Stage Kidney Disease (ESKD) [1–5]. During the past decade and partly fuelled by the KDIGO classification of CKD, many have started questioning whether in the elderly, decreased estimated glomerular filtration rate (eGFR) should really be labelled as a "disease" at all [6]. Simultaneously, a tendency to start renal replacement therapy (RRT) at higher eGFR thresholds, has resulted in a spectacular increase in older people starting dialysis [7, 8]. Strikingly, the higher incidence of RRT in the elderly has been mirrored by an increasing number opting to withdraw from dialysis [9]. Despite large variation in attitudes between regions [10], the idea of conservative care has been gaining traction [11–13]. As a consequence, both nephrologists and patients are currently struggling with how to approach advanced CKD.

A thematic analysis of disease trajectory experiences in elderly patients diagnosed with CKD revealed different themes that should be addressed to improve the care for this population: patients were shocked with their being labelled as having a serious disease, and were anxious about their prognosis; nephrologists felt uncertain about what and how to explain the complexity of the condition, and how to predict and steer future events; patients were...
Shared decision making

Over the last years, shared decision making has been forwarded as an important instrument to improve quality of care [18]. In contrast to the conventional paternalistic approach to medical decision making in which the physician decides what is best for the patient, shared decision making tries to involve the opinion, values and expectations of the patient in the process. For it to be effective, three steps are essential (Figure 1). The first step encompasses informing the patient about the different available options. The expected or most likely outcomes and eventual dangers of these different options should be clearly explained. In the setting of the elderly with advanced CKD, prediction of the outcomes ‘mortality’ and ‘progression to ESKD’ is prime for informing the process of shared decision making. During the past decade, there have been several attempts at developing risk prediction models, combining multiple demographic (e.g. age, sex) and clinical characteristics (e.g. medical history, physical examination results) to estimate individual risk of for each of these situations in the elderly (see below).

Several aspects might jeopardize information transfer. First, the correct information on the expected outcome of the different options might simply not be available, or might not be available for the specific population the patient belongs to (generalizability, external validity). This might be especially problematic in elderly patients with advanced CKD, as both elderly and patients with advanced CKD are mostly excluded from studies, and trials in these people are scant. Unfortunately, data from the general population, or even from the elderly without CKD or the non-elderly with CKD, are not readily translated to the elderly with CKD. Applying prognostic models developed in the general population to elderly patients with advanced CKD may result in overoptimistic prognosis, and pointless technical investigations and care.

Second, the information can be transferred in a biased, non-neutral way. Some treatment routines are so embedded in the structure and paths of care that everybody accepts them as the only possible way, leaving no room for alternatives. Accordingly, information on the alternative options is coloured by non-verbal (or even verbal) signs of disapproval
The need for accurate risk prediction models

Third, the information is often transferred in a way the patient does not understand. It is very difficult to transfer information on probabilities [19] and the uncertainty surrounding them to lay people, and even physicians and healthcare professionals may struggle to understand them [20]. In this era of digitalisation, it is becoming increasingly straightforward to present data in any imaginable graphical format. However, there is limited evidence on which presentations work best, and how these visualisations are processed and understood by the audience [21]. There is some evidence that graphical presentations should be adapted to the target audience, depending upon numeracy and health literacy, and digital tools can help to make these conversions. However, whatever tool or visualisation used, it reduces information from what is already a best guess. The resulting situation is that we are discussing facts that, though objective in themselves, are surrounded by a great degree of uncertainty, leaving ample room for subjective interpretation of the indeterminacy of the future presented. As such, some patients will opt for lower odds than others. Experiments in the field of decision making on dialysis modality highlight that patients tend to be strongly influenced by stories of other patients [22] and far less by the same information provided by a physician.

Fourth, information on outcomes that matter to the patient might not be available, whereas ample information is provided on outcomes that do not matter to the patient [23]. Recently, the SONG project tried to establish a core outcome set for patients on haemodialysis [24]. SONG clarified that most studies use outcomes that are not patient relevant, whereas outcomes that do have value are rarely studied. In 2011, a comparable initiative was taken by the National Institute of Aging. They proposed that studies in this domain should focus on outcomes including measures of pain, fatigue, physical and mental functioning, social roles, daily activities, disease burden, and caregiver burden [25].

A next step in the shared decision making process is the deliberation of the available information. In this step, the physician needs to explore the patients’ wishes, expectations, and values and elicit opinion on different potential scenarios. In this stage, empathy or the skill to view the situation through the eyes of the patient, is a necessary property for the medical team. The deliberation can be flawed if sufficient empathy is lacking, and result in the physician’s rather than the patient’s wish is followed. Empathy can be reduced through distance, either factual – substantial difference in age or different social backgrounds – or created, e.g. by a tendency to see the patient as an object or a case rather than as a person with distinct values and experiences. Empathy can also be endangered by authority, which not only comprises the supervisor or hospital board expecting the patient to start dialysis, but also

**Figure 1 – Shared decision making in steps**

- **1. Inform**
  - Clarify options
  - Describe expected outcomes

- **2. Deliberate**
  - Value clarification
  - Elicit patient preference
  - Suggest, cross-feedback solutions that fit values and preferences

- **3. Decide**

**Problems/Pitfalls**
- Information not available
  - In general
  - For this specific patient population (generalizability/external validity)
  - For the outcome valued by the patient
- Too much information for outcomes not valued by the patient
- Information presented biasedly
- Information presented non-comprehensively

- Bias by your own beliefs and values
- Lack of or insufficient empathy
- Distance
  - Factual distance: differences in age, gender, social circumstances, intellectual capacity...
  - Created distance: objectification, approaching the disease not the patient, hiding behind technical knowledge and skills
- Authority: hospital director, supervisor, guidelines, litigation risk
- Prejudice
- Denial: avoid to bring certain aspects in the discussion (e.g. death)
- Too much identification
- Risk-averse behavior
- Failing to see trade-offs and alternatives
existing guidelines, or fear for litigation. Furthermore, prejudice and denial can compromise open and empathic discussion with the patient. Many physicians find it a major challenge to discuss death with their patients, and will therefore try to avoid these discussions by denying or minimizing risk of death [26], so that patients are presented with unrealistic perspectives. The last step is the actual decision making. Studies indicate that most patients want to be informed about the different options available, but that most will state in one form or another, that the physician should take the decision for them (doctor, what would you do). As discussed above, this does not imply that physicians can simply do as they please, but rather that they should take into account patients’ desires and values, and use their clinical expertise to propose the solution or treatment most likely to result in an outcome the patient desires. This process requires not only empathy, but also insight in how available evidence is applicable to the particular situation of the individual patient, taking into account known comorbidities. It is flabbergasting to realize that for most provided guidance, there is a complete lack of external validity of the underlying evidence for most subpopulations the guideline refers to [27].

### Risk prediction models: definition and assessment pitfalls

In a model of shared decision making, accurate and unbiased information on the fate of the individual is essential. Risk prediction models aim to objectively predict the risk of a future outcome, e.g. mortality or ESKD, based on a set of variables available at the time the prediction is made. Essentially, each variable is awarded a weight – or coefficient – and combined in a mathematical rule to predict an outcome of interest. Similar to weather-forecasts though, it is not because intricate models exist, that they produce reliable estimates of what will happen in reality.

Evaluating the quality, generalizability and utility of risk prediction models poses certain challenges that are both interesting from a methodological point of view and crucial to understand for the clinician wanting to use such models for informing their clinical practice. Some of the methods for assessing model performance can be quite daunting and lead to misinterpretation or overly confident conclusions around accuracy, reliability and generalizability of the predictions (Table 1).

<table>
<thead>
<tr>
<th>Model assessment pitfalls</th>
<th>Limitations</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model only internal validated</td>
<td>Produces overly optimistic view of model performance</td>
<td>External validation in cohorts similar to target population</td>
</tr>
<tr>
<td>C-statistic as measure of discrimination</td>
<td>No practical meaning</td>
<td>Positive predictive values</td>
</tr>
<tr>
<td></td>
<td>No differential weighting of misclassification errors</td>
<td>Negative predictive values</td>
</tr>
<tr>
<td></td>
<td>Heavily dependent on risk factor distribution</td>
<td></td>
</tr>
<tr>
<td>Vague description and dichotomized risk predictors</td>
<td>Classification errors</td>
<td>Clarity of description</td>
</tr>
<tr>
<td></td>
<td>Information reduction and increase in unmeasured variability</td>
<td>Layered risk predictors</td>
</tr>
<tr>
<td>Risk scores with coefficients rounded to nearest integer</td>
<td>Information reduction and unmeasured variability</td>
<td>Risk calculators, apps, integration in electronic health record</td>
</tr>
</tbody>
</table>

1/First of all, a model needs to be tested in a group of people that was not used to develop the model, it needs to be externally validated [28]. Why is that? Well of course, developing a model means mimicking the data as much as reasonably possible and so, often the resulting model will be reasonably good at predicting whatever it is that we want it to predict. As a consequence, conclusions based on performance measures calculated in the same cohort as the one that was used to develop the model, will necessarily produce overly optimistic
conclusions of a model’s accuracy [29]. Ideally, validation is done by investigators who were not involved in the model development process.

2/ The performance of a risk prediction model is commonly assessed by testing its calibration – the agreement of observed and predicted event rates – and discrimination – the ability to distinguish individuals who will develop the outcome of interest from those who will not. Investigators often use the C-statistic as a global measure of model discrimination, ranging from 0.5 (random concordance) to 1 (perfect concordance). The C-statistic can be seen as the area under the curve of the receiver operating characteristic curve, be it with several important limitations [29].

For one thing, it is a single number that does not really have a practical interpretation. Sure, if it equals 1, the model is perfect, and if it equals 0.5, throw it in the bin; but for any number in between, it has no practical meaning attached to it. It does not convey the implications of the misclassification errors that can occur (predicting an individual who experiences an event to be at low risk; predicting an individual who does not experience an event to be at high risk) [29]. For that we need positive and negative predictive values, which do have a direct clinical meaning, but sadly these are seldom reported.

Secondly, the value of the C statistic depends not only on the model being assessed, but also on the distribution of risk factors in the sample to which it is applied. For example, if eGFR is an important risk factor, the same model can appear to perform much better when applied to a cohort with a wide eGFR range than when it is applied to a cohort with a narrow eGFR range. Finally, the C statistic is only a measure of discrimination, so it provides no information regarding whether the overall magnitude of risk is predicted accurately.

For that we need to look at calibration measures, which assess how accurately the model’s predictions match overall observed event rates. In other words, we look at the agreement between what we predict and what we observe. Without going into detail, we can safely state that sadly again, calibration measures are often omitted [30, 31].

3/ For a model to be useful in practice, it needs to include variables that are readily available, well-defined and measurable. Clear definitions of individual risk predictors are necessary to ensure inter-rater reliability. The presence or absence of "diabetes", for example, can be interpreted differently by different raters. Does it apply to everyone meeting international criteria, even if transient or perfectly controlled with limited diet restrictions; or is it limited to patients treated with insulin? The same problem arises for many commonly used predictors, such as cardiovascular disease, peripheral vascular disease, cancer or chronic lung disease: if not explicitly described how these categorical variables should be measured, they can induce substantial variation in scoring, and thus importantly influence model performance [32].

4/ For a model to be applicable in practice, it needs to come with an actual equation that allows straightforward calculation of an individual’s absolute risk of the outcome of interest. To this day, researchers often still opt to create risk scores, which basically transform the model parameters to integers that can be summed to derive a global risk prediction for that individual patient. A classic example we are all very familiar with is the CHADS-Vasc score, which allocates points for age, hypertension, diabetes etc., and relates that sum to an absolute annual risk of stroke in atrial fibrillation. With smartphones being in everyone’s pockets and emergence of companies specializing in developing apps for risk calculation, there seems to be increasingly less virtue in doing that. Instead of simplifying models to allow risk calculation without computer assistance, attention is probably better refocussed to model visualisation to boost uptake and efficient communication of their results [21].

Available risk prediction scoring systems for elderly with advanced CKD

Progression to ESKD

Several prospective [33–35] and retrospective [4, 5, 36–38] cohort studies aimed to develop risk prediction scores based on identified risk factors for progression to ESKD. The Kidney Failure Risk Equation (KFRE) initiative analysed data from Canadian adults with eGFR 10–59 ml/min/1.73m² to develop the KFRE equation to predict the risk of ESKD at 2 and 5
years [38]. Using 8 variables (age, sex, eGFR, albuminuria, serum calcium, serum phosphate, serum bicarbonate, and serum albumin) this score proved to have good discrimination capacity, both in development (C statistic = 0.92) and in the validation cohorts (C statistic = 0.84), and reducing the number of variables further down to 4 easily available parameters (age, gender, eGFR and albuminuria) did not substantially alter the discriminative capacity performed similarly (C statistic = 0.91 and 0.84 in development and validation cohorts, respectively) [39]. An external validation was carried out in a Dutch cohort with stage 3–5 CKD, demonstrating that the scores performed well to predict 5-year risk (C statistic 0.89 and 0.88, respectively) and also had a good calibration (difference between predicted and observed risk 4.0% and 7.1%, respectively). In an external dataset with over 700,000 patients, the 4-variable KFRE achieved excellent discrimination (poored C statistic 0.90 at 2 years and 0.88 at 5 years), although the KFREs tended to overestimate risk in some non-North American cohorts. Addition of a calibration factor improved calibration in 12/15 and 10/13 non-North American cohorts at 2 and 5 years, respectively [40]. These data seem to indicate that the 4 variable KFRE (with the use of a calibration factor for non-American cohorts) is suitable to estimate the risk for evolution to ESKD in this population.

**Mortality risk in CKD**

Since 2012, two high quality systematic reviews, one including models predicting death in elderly people [41] and one predicting death in people with CKD [42] have been published on the topic. Starting from the search strategies of these papers, we identified 24 publications including 31 risk prediction models, of which 15 models target elderly people in general [43–55], 4 elderly people with CKD 3–5 [5, 56, 57], and 12 elderly people with end-stage kidney disease (ESKD) [58–63]. Only three models were developed or validated in Western Europe. The most commonly included final predictors of death were age, sex, functional status, heart failure, malignancy and diabetes. Although most models included parameters of frailty, only one model was specifically developed within a frail elderly patient group [43]. As a consequence, it would be safe to consider using an additional scoring system for frailty in patients with a low predicted risk for mortality by any of these scores.

Another caveat is that external validation was mostly not available, and as far as it was, it was mostly done by the same investigators that had developed the model, and in patients very similar to the ones included in the development cohort. In addition, presence of comorbidities was based either on self-report or on coding within administrative databases. Both methods can induce misclassification as criteria might not be clear or well described. This can substantially reduce predictive performance, especially upon generalisation to patient groups external to the ones used for model development. In general, model performance was moderate at best, with only 1 model achieving a c-statistic of > 0.8 [49], and confidence intervals were generally not provided. The Bansal risk prediction model predicts the absolute probability of death within five years for older people with CKD stage 3 through 5 not yet treated with dialysis, provides measures of predictive performance and was externally validated in a large cohort of representative patients, except that the overwhelming majority was able to live independently [56]. The model has a reasonable calibration and model discrimination in both the development (0.72; 95% CI 0.68 to 0.74) and validation cohort (0.69; 95% CI 0.64 to 0.74). As the validation cohort might not be representative for cohorts containing frail patients, and as it has been well established that frailty is a prevalent condition in patients with advanced CKD (eGFR<45ml/min) [64], it is absolutely mandatory to combine the Bansal score with a score for frailty when Bansal score is low, as in this setting mortality risk will be governed by frailty rather than by traditional risk factors. However, it has been advocated that frailty is an additional risk factor for mortality, on top and independent of other traditional risk factors [65]. As such, a high predicted mortality with the Bansal score will deliver a reliable result even in a frail patient.

**Mortality risk in ESKD**

One risk prediction model based on the REIN-registry data estimates risk of death at
three months in older people with ESKD starting with dialysis [60]. The cohort is representative for elderly patients starting dialysis, both in terms of age (at least 75 years old, with one in five > 85 years) and in terms of comorbidity (heart failure 33% and peripheral vascular disease 25%). This risk prediction model includes 9 easily available predictors: age, sex, history of congestive heart failure, peripheral vascular disease, arrhythmia, cancer, severe behavioural disorder, mobility and baseline serum albumin concentration. The rate of death in the validation cohort increased with the score, indicating good calibration, but discrimination was moderate with a c-statistic in the internal validation cohort of 0.75 (95% CI: 0.74–0.76). The model was further externally validated in a Flemish cohort [39], although the investigators slightly modified the score. Another risk prediction score based on the REIN-cohort data estimates risk of death at six months in older people with ESKD starting with dialysis [59]. The model was further externally validated in an American cohort [58] although again, investigators modified the score [58].

Floege and co-workers also developed a risk prediction model predicting mortality in patients starting dialysis based on the Framingham study model [66]. This model was then validated in an external cohort of the Dialysis Outcomes and Practices Patterns (DOPPS), showing a moderate discrimination (c statistic of 0.68 to 0.79 depending upon geographic location). Although the score includes age, it has not been developed or validated in a cohort of elderly dialysis patients (mean age 64 ± 14 years), and the cohort did also not include peritoneal dialysis patients. Furthermore, the development cohort includes only patients who survived the first 3 months, whereas the validation cohort of DOPPS includes mainly prevalent patients. Both attributes make that the score is likely to be not very representative for the dilemma whether or not to start dialysis in the frail elderly, where exactly this risk of short term mortality during the initiation phase of dialysis is what needs to be predicted.

Conclusions
In conclusion, reliable, externally validated risk prediction models for progression of CKD to ESKD or mortality in frail elderly with or without CKD are necessary to inform shared decision making in the trajectory of the elderly with advanced CKD, but available models are scant. Physicians need to understand the limitations of these models so that they can be used appropriately. Next to understanding the models themselves, healthcare workers need to translate the available information to patients, and that in a way the patient can understand, and use the information to choose a trajectory of care most likely to achieve his/her goals and expectations, taking into account his individual needs. Rather than developing new models in a search for more sophisticated statistical models, we emphasize the importance of external validation by different investigators of those models in both frail and non-frail elderly patients to test their performance and applicability. In addition, more effort must go to developing strategies for translating the information such that it becomes digestible and understandable for all involved.

REFERENCES


Резиме

ПОТРЕБА ОД ПРЕЦИЗНИ МОДЕЛИ ЗА ПРЕДВИДУВАЊЕ РИЗИК, ЗАЕДНИЧКО ДОНЕСУВАЊЕ ОДЛУКИ И ПЛАНИРАЊЕ НЕГА НА ПОСТАРИ ЛИЦА СО НАПРЕДНА ХРОНИЧНА БУБРЕЖНА БОЛЕСТ

Маријке Стикерс, Еви В Наглер, Вим Ван Бисен
Оддел за нефрологија, Универзитетска болница во Гент, Гент, Белгија

Како што стареат луѓето, хроничната бубрежна болест станува се почеста, но ретко доведува до краен стадиум на бубрежна болест. Кога се случи тоа, изборот помеѓу дијализа и конзервативна нега може да биде ефикасно заедничкото донесување одлука. Моралите модели за предвидување ризик може да помогнат ако се надворешно потврдени, добро калибрирани и дискриминативни; ако вклучуваат недвосмислени и мерливи променливи; и ако доаѓаат со применливи равенки или резултати. Недоволни се сигурните, надворешно потврдени модели за предвидување ризик од прогресија на хроничната бубрежна болест до краен стадиум на бубрежна болест или смртност кај немошни стари лица со или без хронична бубрежна болест. Во рамки на овој труд, разгледуваме голем број надежни модели, истакнувајќи ги предностите и ограничувањата што треба да ги разберат лекарите за да ги користат разумно и да ја истакнуваат потребата од надворешна валидација преку нов развој за натамошно унапредување на полето.

Ключни зборови: прогноза, пропорционални модели на ризик, логистички модели, возрасни лица, бубрежна инсуфициенција, хронична