

University of Groningen

Prediction and monitoring of chronic kidney disease

Schutte, Elise

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2017

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Schutte, E. (2017). *Prediction and monitoring of chronic kidney disease*. [Thesis fully internal (DIV), University of Groningen]. Rijksuniversiteit Groningen.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

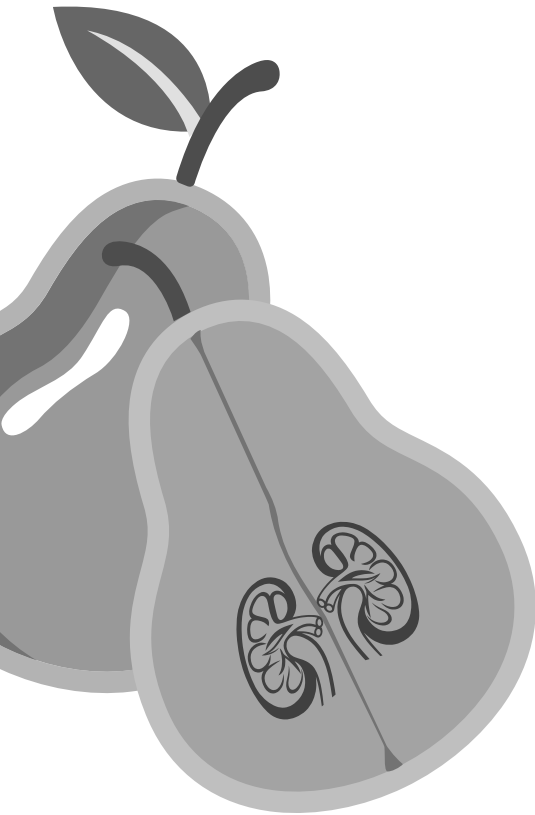
Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Chapter 4

Will the future lie in multitude? A critical appraisal of biomarker panel studies on prediction of diabetic kidney disease progression



Elise Schutte
Ron T. Gansevoort
Jacqueline Benner
Helen L. Lutgers
Hiddo J. Lambers Heerspink

Nephrol Dial Transplant 30 Suppl 4: iv96-104, 2015.

Abstract

Diabetic kidney disease is diagnosed and staged by albuminuria and eGFR. Although albuminuria has strong predictive power for renal function decline, there is still variability in the rate of renal disease progression across individuals that is not fully captured by the level of albuminuria. Therefore, research focuses on discovering and validating additional biomarkers that improve risk stratification for future renal function decline and end-stage renal disease in patients with diabetes, on top of established biomarkers. Most studies address the value of single biomarkers to predict progressive renal disease and aim to understand the mechanisms that underlie accelerated renal function decline. Since diabetic kidney disease is a disease encompassing several pathophysiological processes, a combination of biomarkers may be more likely to improve risk prediction than a single biomarker. In this review we provide an overview of studies on the use of multiple biomarkers and biomarker panels, appraise their study design, discuss methodological pitfalls, and make recommendations for future biomarker panel studies.

Introduction

Diabetic kidney disease is diagnosed, monitored for progression, and staged for risk of progression by a combination of biomarkers. The National Institutes of Health biomarker definition working group defines a biomarker as “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention”.¹ The key biomarkers that are used in diabetic kidney disease are albuminuria and eGFR.² The Kidney Disease Outcomes Quality Initiative (KDOQI) clinical practice guideline advises to measure these biomarkers at least yearly in all patients with diabetes, starting 5 years after diagnosis in patients with type 1 diabetes and at diagnosis in patients with type 2 diabetes.²

From a preventive medicine point of view, there is a need for novel biomarkers in the early stages of diabetic kidney disease, because the earlier renoprotective treatment is started, the better long-term outcome will be. One challenge of developing novel biomarkers is to identify which patients are at high enough risk during early stages of disease to warrant the start of treatment, that will inevitably be associated with adverse events. In that respect albuminuria is of higher interest than eGFR, because by definition, low eGFR identifies only patients at late stage disease. Fortunately, subtle increases in albuminuria are strongly predictive for progression of diabetic kidney disease, even in the early phase of the disease when eGFR is still normal.³ The predictive performance of albuminuria has been corroborated in a vast number of independent diabetic populations. Although albuminuria has strong predictive power, there still is variability observed in the rate of renal disease progression that is not fully captured by the level of albuminuria. Despite extensive research in the last 50 years, it is disappointing to notice that no new single biomarker has been discovered that unequivocally outperforms albuminuria or adds to albuminuria in predicting kidney disease progression.⁴ Although some promising biomarkers have been discovered, such as TNF-Receptor-1, most biomarkers failed to replace albuminuria or add significantly to albuminuria as predictor of diabetic kidney disease progression.

Since there are multiple underlying pathophysiological processes involved in diabetic kidney disease, such as inflammation, angiogenesis and fibrosis, it may be difficult for a single biomarker to predict the progression of diabetic kidney disease. Therefore it seems more logical to develop panels of biomarkers that capture several pathophysiological processes to improve prediction of diabetic kidney disease progression. These panels can exist as a combination of biomarkers that have been identified in targeted single biomarker studies or biomarkers that have been discovered through hypothesis-free research methods, for example, by means of proteomic or metabolic profiling. We refer to the review by Pena et al. elsewhere in this issue of NDT for a detailed description of proteomics and other new high-throughput biomarker discovery methods.⁵

The translation of a biomarker or a combination of biomarkers from discovery to clinical practice is a process full of pitfalls and limitations. Before a biomarker can be used in clinical practice, it needs to be extensively validated in large studies to assess accuracy, reproducibility, sensitivity, and specificity. When conducting these studies, key issues such as trial design and methodology should be addressed. Therefore, the aim of this article is to review the available literature on biomarker panels in diabetic kidney disease. We will discuss methodological pitfalls and provide recommendations for future biomarker panel research.

Overview of biomarkers used for prediction of diabetic kidney disease progression

Although numerous studies have been conducted that investigate biomarker panels in chronic kidney disease the literature on biomarker panels for predicting specifically diabetic kidney disease progression is scarce. We performed a systematic literature search for studies on biomarker panels, or studies investigating combinations of multiple biomarkers in diabetic kidney disease, using the search strategy and inclusion criteria depicted in Box 1.

Box 1 Search strategy

Pubmed search strings: Combinations of the terms diabetic nephropathy; diabetic kidney disease; multiple biomarkers; biomarker; markers; biomarker panel; prediction; predictor; predictors; renal function decline

Inclusion criteria:

- Population: adult patients with diabetes (both type 1 and 2)
- Endpoints: incidence or progression of diabetic kidney disease
- Multiple markers (≥ 2) had to be investigated
- Analysis of markers as a panel in (adjusted) statistical models (instead of separate analyses for single biomarkers)

No. of articles found: 1440 total

Articles fulfilling inclusion criteria: 3

Articles identified by cross-references and expert advice: 6

Studies using -omics methods were excluded since these methods will be discussed elsewhere in this issue of NDT by Pena et al.⁵

Table 1 provides an overview of the studies that were identified and summarizes the biomarker panels that were investigated and methodology that was used. We identified 9 studies.⁶⁻¹⁴ The design of these studies varied from observational cohorts (n=5) to post hoc analyses of randomized controlled trials (n=3) and case-control studies (n=1). Sample size across the studies ranged from 82 to more than 10,000 patients, and the number of candidate biomarkers varied from 2 to 28. The biomarkers that were studied varied from kidney specific biomarkers (e.g. Tamm-Horsfall protein) to systemic biomarkers (e.g. hs-CRP). Urinary biomarkers were investigated in three studies,^{8,10,12} serum or plasma biomarkers in four studies,^{6,9,11,13,14} and one study combined urine and serum biomarkers.⁷ The endpoints that were chosen to indicate renal disease progression were hard renal endpoints (including end-stage renal disease (ESRD),⁹⁻¹¹ and/or a combination of ESRD and death^{7,11}), as well as surrogate endpoints (including eGFR decline⁶⁻⁸ and the development of micro- or macroalbuminuria^{10,12,14}). Follow up time in the studies ranged from 2 to 6 years.

Table 1 Overview of the 9 selected studies investigating the predictive value of biomarker panels for diabetic kidney disease progression

Study	Design	Sample size and diabetes type	Definition of diabetic kidney disease	Number of candidate biomarkers	Pathophysiological domains represented by biomarkers	Matrix
Pena et al. 2015	Observational cohort study	82 DM2	NR	28	Inflammation, fibrosis, angiogenesis, endothelial dysfunction, mineral metabolism, lipid metabolism, glomerular damage	Serum
Agarwal et al. 2014	Case-control study	67 + 20 DM2	Clinical criteria	24	Inflammation, fibrosis, angiogenesis, glomerular injury, mineral metabolism, tubulointerstitial injury	Urine/plasma
Verhave et al. 2013	Observational cohort study	83 DM1+2	Proteinuria >0.5 g/day	7	Inflammation, fibrosis	Urine
Wong et al. 2013	Case-control study	281 DM2	NR	2	Fibrosis (TGF- β , BMP)	Plasma
Schatzler et al. 2012	Observational cohort study	652 DM1	NR	4	Inflammation, angiogenesis, endothelial dysfunction, tubular injury	Urine
Desai et al. 2011	Prospective cohort within RCT	1000 DM2	NR	2	Cardiac markers (hsTNT, NT-proBNP)	Serum
Kern et al. 2010	Nested case-control	91 + 178 DM1	AER >40mg/day or AER >300 mg/day	4	Inflammation, tubular injury, AGEs	Urine
Astrup et al	Observational cohort study	199 + 192 DM1	AER >300mg/24h +retinopathy	7	Inflammation, endothelial dysfunction	Plasma

Table 1 (Continued)

Study	Design	Sample size and diabetes type	Definition of diabetic kidney disease	Number of candidate biomarkers	Pathophysiological domains represented by biomarkers	Matrix
Persson et al. 2008	Post-hoc analysis of RCT	269 DM2	AER >200 µg/min + 3% increase in AER from baseline	9	Inflammation, endothelial dysfunction	Plasma
Pena et al. 2015	Age, sex, eGFR, UACR, SBP, oral diabetic medication, smoking	eGFR decline	4.0	78±23	AUC	
Agarwal et al. 2014	Age, eGFR, UACR	eGFR decline, ESRD/Death	1.8	43±13	NR	
Verhave et al. 2013	Renal function decline p/year, proteinuria, SBP, DBP, hba1c, oral hypoglycemic drug/insulin, statin use, RASB, treatment	eGFR decline	2.1	25±9	NRI	
Wong et al. 2013	Age, sex, eGFR, UACR, HbA1c, diabetes duration, BMI, history of macrovascular or microvascular disease	Major renal endpoint (DSCR, ESRD, renal death)	5	71 ±16	AUC	
Schatzler et al. 2012	Age, AER, SBP, HbA1c, WHR, diabetes duration, BMI, smoker, antihypertensive treatment, anti-lipid Treatment, uric acid, cystatin c, cholesterol, HDL	Macroalbuminuria, ESRD, ESRD + macroalbuminuria	6	>125	AUC	
Desai et al. 2011	Age, eGFR, UPCR, insulin use, bmi, race, SUN, albumin, hemoglobin, ferritin, CRP, history of AKI, prior stroke/pad/hf, cardiac arrhythmia	ESRD, Death/ESRD	3.5	<40*	NRI	

Table 1 (Continued)

Study	Design	Sample size and diabetes type	Definition of diabetic kidney disease	Number of candidate biomarkers	Pathophysiological domains represented	Matrix
Kern et al. 2010	Age, sex, baseline eGFR, MAP, baseline HbA1c, average follow-up HbA1c, diabetes duration	Macroalbuminuria, microalbuminuria	NA	>125*	NR	
Astrup et al	Age, sex, UAE, SBP, smoking, A1c, total cholesterol, heart rate variation	Death, cardiovascular death/morbidity	10	74*	NR	
Persson et al. 2008	Age, sex, UAER, MABP, HbA1c, treatment	Macroalbuminuria	2	80†	AUC	

Abbreviations are: UACR, urine albumin-creatinine ratio; SBP, systolic blood pressure; eGFR, estimated glomerular filtration rate; DBP, diastolic blood pressure; RASB, renin angiotensin system blocking medication use; WHR, waist hip ratio; BMI, body mass index; HDL, high-density lipoprotein; AER, albumin excretion rate; SUN, serum urea nitrogen; UPCR, urine protein creatinine ratio; PAD, peripheral artery disease; HF, heart failure; CRP, c-reactive protein; AKI, acute kidney injury; MAP, mean arterial pressure; AGEs, advanced glycation end product; RCT, randomized controlled trial; NA, not applicable; NR, not reported; AUC, area under the curve
 *: overall eGFR was not reported, the study reported mean eGFR for subgroups. The table represents an overall range of mean eGFR for the subgroups.
 †: eGFR was not reported. The here reported eGFR was calculated using the CKD-EPI calculation for the overall mean creatinine (90 $\mu\text{mol/L}$) for males aged 57.2 years (the overall mean age).

Points of interest in biomarker panel studies

As the identified studies varied to a large extent in their design, choice of biomarker panels, and methodology of statistical analysis, we performed an analysis according to our prior defined criteria which are presented in Box 2. Furthermore, we assessed the phase of biomarker development for each study, as defined by the American Heart Association in 2009 (Figure 1).¹⁵

Box 2 Key points for conducting biomarker studies

Phase of development

- Phase 1 - 5 (figure 1)

Population

- Early or late-stage disease

Endpoints

- Surrogate endpoints
- Hard endpoints

Measurement of biomarkers

- Selection of matrix (blood, urine)
- Handling, freezing and storage of samples before analysis
- Assay characteristics
- Cut-off points for biomarker

Statistical analysis

Sample size and number of events

- Adjustment for covariates (etiologic or predictive)
- Discrimination (AUC, NRI, IDI)
- Calibration (observed and expected event rates)

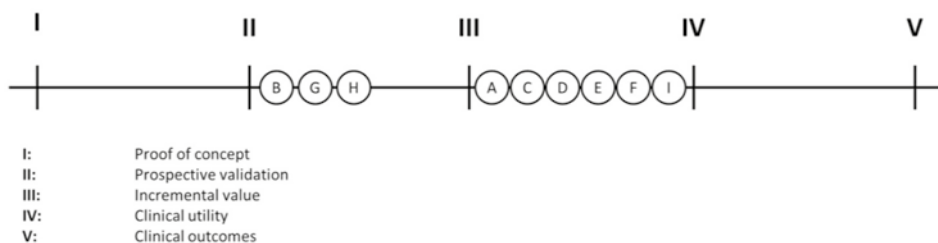


Figure 1 Phases of biomarker evaluation

Visualization of the biomarker evaluation progress according to the American Heart Association, consisting of five phases¹⁵. The nine studies included in this review are represented by letters A-H (according to year of publication, with A being the most recent study), positioned on the axis according to evaluation phase. A, Pena et al 2015; B, Agarwal et al. 2014; C, Verhave et al. 2013; D, Wong et al. 2013; E, Schlatzer et al. 2012; F, Desai et al. 2011; G, Kern et al. 2010; H, Astrup et al. 2008; I, Persson et al. 2008.

Phase of development

Prior to 2009, no formal guidelines for the development and validation of new biomarkers existed. Hlatky et al. proposed a framework for the development of biomarkers in collaboration with the American Heart Association in 2009.¹⁵ This biomarker development framework consists of five phases. First, the proof of concept phase tests whether biomarker levels differ between subjects with and without the outcome in a cross-sectional study. Second, the validation phase investigates the association of the new biomarker with hard endpoints in a prospective cohort or prospective case-control study. Third, the assessment of the incremental value phase tests whether the new biomarker adds predictive value on top of standard risk markers. Fourth, the clinical utility phase assesses whether the biomarker changes the predicted risk sufficiently enough to warrant initiation or change of therapy. Fifth and final phase, the validation phase consists of a randomized controlled clinical trial in which a biomarker panel is measured in patients, who are then divided into two strata based on the biomarker panel results and subsequently randomly assigned to treatment or to standard of care, to assess whether treatment guided by the new biomarker panel improves clinical outcomes.

The nine studies reviewed here represent phase 2 (n=3) or phase 3 studies (n=6) (Figure 1). Thus, it can be concluded that biomarker panels for prediction of diabetic kidney disease progression are still in the initial phases of development as no external validation studies have (yet) been performed. Moreover, using biomarker panels to assess treatment consequences is lacking, let alone the initiation of randomized controlled clinical trials. Implementation of biomarker panels in clinical care requires that all steps of the biomarker development process are followed. Consequently, there is a long way to go before biomarker panels can be used in clinical practice to guide treatment in patients with diabetic kidney disease.

Population

All studies included rather homogenous cohorts with respect to diabetes status, gender, race, and severity of diabetic kidney disease. Three studies exclusively investigated patients with type 1 diabetes.^{10,12,13} These patients were on average young and had early stage kidney disease. The other six studies investigated patients with type 2 diabetes. Four studies measured the biomarkers in a predominantly male cohort,^{7-9,14} and all study populations consisted almost entirely of Caucasians.

Only four studies reported the exact definition they used for the diagnosis of diabetic kidney disease (either for inclusion or as endpoint).^{8,12-14} Diabetic kidney disease is present in patients with macroalbuminuria (>300 mg/g creatinine in spot urine) irrespective of the eGFR level, and possible in patients with microalbuminuria (30-300 mg/g creatinine in spot urine) and eGFR >30 ml/min/1.73m², according to the KDOQI guideline for diabetes and CKD.² Different definitions for diabetic kidney disease were used in the studies included in our review, yet we have to acknowledge that the latest KDOQI guideline became available after these studies were initiated. For future studies, it is important that all studies use the same definition of diabetic kidney disease in order to facilitate comparison of study populations and results across studies.

The studies included patients at either early or late stage of diabetic kidney disease. These relatively homogeneous cohorts reduce the external validity for study results to be extrapolated to

the general diabetic population. Therefore it is recommended that the ideal cohort includes a large number of individuals with sufficient variation in age, gender, race, and in renal risk markers to allow well powered subgroup analyses. Within the SysKid (Systems Biology towards Novel Chronic Kidney Disease Diagnosis and Treatment) program, a European research project on biomarkers for kidney disease,¹⁶ a large scale biomarker panel validation study is conducted in such a heterogeneous cohort of type 2 diabetic kidney disease patients, including subjects with both early and late stage of disease.

Of note, two of the included studies were post hoc analyses of clinical trials.^{11,14} When using a clinical trial cohort, treatment effects on clinical outcome have to be taken into account. The study drug can affect a clinical outcome, which will likely alter the association between baseline biomarker levels and long-term outcome. In addition, clinical trials are not representative of clinical practice as patients are more strictly followed, leading to higher compliance and better outcomes, even with standard care, than in the general population. Therefore, information on study treatment as well as changes in (standard) treatment during follow-up should be taken into account. Failure to collect, analyse, and report this information may lead to reduced predictive performance of a candidate biomarker.

Finally, it is important that all study participants come from the same population to prevent selection bias. Selection bias in an experimental study refers to the situation when the selection of patients is not sufficiently random to draw a general conclusion. For example, selection bias may arise if one aims to develop a biomarker panel for prediction of renal endpoints for the general type 2 diabetic population, but one creates a study cohort including only patients with diabetic kidney disease. This can lead to the question whether the biomarker panel is generalizable to the general type 2 diabetic population or only for patients with established diabetic kidney disease.

Endpoints

According to the biomarker evaluation framework, biomarker validation should preferably involve hard outcomes. The hard outcomes to study diabetic kidney disease progression that have been accepted by the United States Food and Drug Administration (FDA) and the European Medicines Agency (EMA) include the incidence of ESRD, defined as the need for chronic dialysis or kidney transplantation or renal death. Three studies used hard endpoints only,^{9,11,13} two used a combination or composite of hard and surrogate endpoints^{7,10} and four exclusively used surrogate endpoints^{6,8,12,14} (Table 1). The hard endpoints were ESRD, death, a composite of ESRD and death, a composite of ESRD or doubling of serum creatinine and cardiovascular mortality or morbidity. The surrogate endpoints that were used were onset of micro- and/or macroalbuminuria and eGFR decline.

Only half of the studies used hard endpoints as outcome measures. This is explained by the fact that kidney disease is in general slowly progressive and that hard endpoints take a long time to occur. Therefore, to accrue a sufficient number of hard endpoints within an acceptable period of time, studies are required that typically enrol large numbers of patients at later stages of disease. Indeed, the studies in our review that used hard endpoints were larger in size and all enrolled patients with severe diabetic kidney disease (impaired eGFR and high albuminuria). This limits the generalizability of the studies' results since the performance of the biomarker panels in early stage

of disease cannot be assessed. As discussed above, this is unfortunate, because from a preventive medicine point of view, there is especially a need for novel biomarkers in the early stages of diabetic kidney disease. For this reason there is an interest in developing alternative, surrogate endpoints that occur more frequently and earlier in the course of the disease. The FDA and EMA have already accepted doubling of serum creatinine (corresponding to an approximately 57% decline in estimated GFR using the CKD-EPI equation) as a surrogate end point. However, this is still an outcome that occurs infrequently and in late-stage disease. Recently, the US National Kidney Foundation and the FDA organized a workshop that explored whether lesser declines in eGFR (i.e. a confirmed 30% or 40% decline in eGFR) may be acceptable as surrogate endpoints for kidney disease progression.¹⁷ These endpoints, as well as the incidence of micro- or macro-albuminuria,¹⁸ are promising candidate surrogate endpoints because they will facilitate renal research and stimulate clinical trials; however, these surrogate endpoints are yet to be thoroughly validated.

Measurement of biomarkers

Choice of matrix (serum, plasma, or urine), sample handling, storage conditions, and assay characteristics are important yet sometimes neglected details when reporting biomarker studies. The nine reviewed studies reported choice of matrix for (a combination of) three reasons: a) based on literature review,^{6-9,12-14} b) accessibility of the matrix¹⁰ and c) a known or hypothesized disease pathway for a biomarker in a specific matrix.¹⁰⁻¹² Ideally, biomarkers are to be analysed in a matrix that is easy to obtain (preferably non-invasive) and the biomarker level in that matrix should represent the disease process that is investigated. However, for many biomarkers it is still unclear whether and where they are produced in the disease process. Future research is needed to determine what matrix is preferred for which biomarkers.

Assay characteristics were reported in all nine reviewed studies. However, none of the studies report the rationale behind the choice of a specific assay, which can have consequences. This is illustrated by a study from the oncology area.¹⁹ In that study a commercially available ELISA was used to validate new biomarkers for pancreatic ductal adenocarcinoma (PDAC). Initially they found that zona pellucida-like domains protein 1 (CUZD1) was a promising new biomarker for PDAC. However, when performing additional laboratory analyses to confirm that the assay had indeed detected CUZD1, the researchers found that their assay was hampered by cross-reactivity and that instead of CUZD1, the assay had detected cancer antigen 125. It is therefore of great importance to report the key features of the assay used, including the lower limit of detection, accuracy (including cross-reactivity), and inter- and intra-assay coefficients of variation.

With respect to sample handling, the effect of frozen storage on the biomarker level is an often overlooked issue, which may have a large impact on study results. It is known that the stability of biomarker levels varies per biomarker and for different storage conditions (e.g. unstable at -20 and stable at -80°C).²⁰ It is therefore recommended to assess the effect of frozen storage on biomarker levels before conducting large scale validation studies using samples that are stored for prolonged period of time. Frozen storage that results in a decrease in average biomarker concentration, as well as an increase in variability in biomarker concentration, will impair the predictive performance of biomarkers, as has been described for urinary albumin.²¹ Sample storage conditions were mentioned

in all but one of the studies included in this review.¹⁴ Only one study reported that levels of urinary marker pentosidine were stable during eight years of storage at -80°C, but specific and detailed information was lacking.¹² The lack of information on the impact of frozen storage on biomarker levels suggests that this important issue may have been negated. Researchers are therefore advised to assess and report stability during frozen storage of biomarkers in the matrix they used.

Statistical analysis

To properly assess the predictive performance of biomarker panels, and especially to minimize the chance of false negative results (type II error), a formal sample size calculation should be conducted before the study is initiated. For such a sample size calculation, the expected predicted power of the biomarker panel is necessary, as well as the between individual variation in the biomarker panel, the correlation with known risk predictors, and the endpoint event rate. None of the included studies performed a sample size calculation, and all studies included relatively small patient cohorts, with the exception of the study of Desai.¹¹

Testing the predictive value of candidate markers in the final biomarker panel is an important topic, but surprisingly no commonly agreed procedure exists as to how this should be performed. In general three strategies are used. First, the predictive value of markers can be tested in a full model approach including all markers. This approach avoids selection bias and provides accurate estimates and standard errors.²² However, selection of all markers is not practical for future clinical care. An alternative approach is a backward elimination selection. In this procedure candidate markers are eliminated from the model if they don't meet a predefined nominal significance level that is often set at 5%. The selection of markers through backward elimination depends, however, on the choice of the significance levels and the size of the population that is studied. A less stringent significance level results in a biomarker panel with more candidate markers, as will be the case in large populations. An alternative to using the backward elimination is the Akaike Information Criteria which is a measure of model fit and includes a penalty of adding additional candidate markers in the model and hence avoids over fitting.²³

The studies that are reviewed were all conducted to assess the predictive performance of the biomarker panels beyond traditional clinical risk markers. To assess the improvement in predictive performance, all renal risk predictors should ideally be taken into account that are used in normal clinical care. In this respect, the quality of the studies has apparently not improved over the last years. A systematic review conducted a couple of years ago concluded that less than half of all studies adjusted for most acknowledged risk factors of diabetic kidney disease progression.²⁴ In the present review, most studies, even the recent ones, did not adjust for all acknowledged renal risk markers.

Prediction models should discriminate between those patients who will, versus those who will, not develop the event of interest. Measures of discrimination are known as the C-statistic, or the area under the receiver operator characteristic curve (AUROC). However, the use of the C-statistic is limited by the difficulty of interpreting the clinical significance of the usually small but statistically significant increase in C-statistic and the direct relation of the increase in C-statistic to the performance of the baseline model.²⁵⁻²⁷ A clinically more useful method is to calculate whether the addition of a new biomarker changes individual risk prediction to such extent that it leads to a

change in the management of the patient (e.g. drug prescription or discontinuation). New metrics such as the net reclassification improvement (NRI) and the integrated discrimination improvement (IDI) have been developed to address this issue. The NRI involves a reclassification table to calculate how many individuals with the event are reclassified to a higher risk category when adding the novel biomarker(s) to the existing risk scores and on the other hand how many individuals without the event are re-classified to a lower risk category. The sum of these net percentages is represented by the overall NRI and its theoretical range lies between -2 and +2. Large positive values indicate that the novel biomarker aids in identifying individuals at risk of the outcome. This enables the clinician to tailor therapy to those individuals at highest risk. On the other hand, negative values indicate that the novel biomarker aids in identifying individuals not at risk and may therefore be useful in preventing overtreatment. In order to determine whether the overall NRI is driven by the positive or negative component, it is recommended to also report the two components of the NRI separately.²⁸ The IDI can be considered an extension of the NRI. The IDI does not include a priori defined clinically meaningful cut-offs for the probability of the outcome (e.g. 5%, 10% or 20%) but integrates the NRI over all possible cut-offs for the probability of the outcome.²⁹ A recent review and clinician's guide of the NRI is provided by Leening et al.²⁸ Most biomarker studies included in our systematic review reported measures of discrimination, either Area Under the ROC Curve, NRI or NRI. However, when the NRI was reported, only the overall NRI was reported and the two components of the NRI (reclassification to higher risk for those with event and reclassification to lower risk category for those without the event) were lacking.

Another important part of validation of biomarker studies is calibration. Calibration is reported as the agreement between observed and predicted endpoints.³⁰ This can be assessed graphically (plotting the predicted versus the observed risk) or with a goodness of fit test, such as the Hosmer-Lemeshow test. In a development cohort, calibration is usually good, therefore reporting of calibration is most important in external validation studies (phase IV and V), which none of the studies reviewed here did.

Although discrimination and calibration are important aspects to verify the utility of the model, these measures do not assess whether clinical decisions improve with the help of the model. A prediction model that aids in clinical decision making requires a cut-off to classify patients in low or high risk groups. The use of a multiple biomarker panel will usually include a risk function, which provides a probability threshold. Classification of patients will thus be based on a decision probability threshold, which is defined as the threshold at which the likelihood of benefit balances the likelihood of harms. For example, a threshold of 4% indicates that start of dialysis of a patient not treated with an angiotensin receptor blocker is 96:4=24 times worse than the complications of hyperkalemia of a patient unnecessarily treated with an angiotensin receptor blocker. Once a threshold is chosen, the sensitivity and specificity can be calculated. The sensitivity and specificity can be calculated for various decision thresholds, and an overall measure of the model's usefulness can be obtained. The disadvantage of this approach is that calculation of sensitivity and specificity at various thresholds negates the relative weight assigned to detecting true disease (true positive; TP) versus over-diagnosing non-disease (false positive; FP). Net Benefit is a novel measure and is defined as $NB = (TP - wFP)/N$ where N is the total number of patients and w is the weight for appropriate

diagnosing (TP) versus over-diagnosing (FP). The additional value of a biomarker panel can be assessed by calculating the difference in net benefit at a certain decision threshold for predictions with and without using the biomarker panel. Finally, the net benefit can be visualized in a decision curve in which the Net Benefit is plotted against the possible risk thresholds for treatment. We refer to Kerr et al. for a more detailed description and examples of Net Benefit analyses.³¹

Ideal study designs

Based on the key points we have discussed from the reviewed studies, we now propose a framework for an ideal study design in biomarker panel validation. A new biomarker panel ideally goes through all five stages of biomarker development, from discovery and proof of concept to addressing clinical utility. This means that a biomarker panel has to be studied in several patient cohorts. Ideally these studies will be performed in a diabetic kidney disease population with varying degree of kidney impairment, to allow for extrapolation of results to the broader type 2 diabetic population. Preferably, hard endpoints should be used, especially in phase III to V studies. However, we encourage the development and validation of surrogate outcomes to allow for prognostic assessment of biomarker panels in earlier stages diabetic kidney disease. The matrix (usually blood or urine) should represent the disease process under investigation, and the available literature on the subject should be described in the report. In addition, it is important to report the assay used along with its key features. Furthermore, when dealing with samples that have been stored for a prolonged period of time, it is advised to assess and report the stability of the biomarkers of interest. A sample size calculation should always be performed before the start of the study to prevent type II errors. For the assessment of predictive power of a biomarker panel, we advocate the use of either the NRI or IDI. We highly encourage the inclusion of all renal risk predictors that are used in daily clinical care in the predictive models, since the new biomarker will be used on top of these clinical risk markers. Furthermore, a phase IV or V study should always report calibration to validate the predictive capabilities of the biomarker panel. Finally, the net benefit score should be calculated and used to translate the predictive power of a biomarker panel into decisions for clinical practice.

Conclusions

In this review we analyzed the methodology of nine studies that investigated biomarker panels for prediction of diabetic kidney disease progression. All reviewed studies represent early phases of the biomarker panel development process. None assessed clinical utility of the investigated biomarker panels or the effect the biomarker panel may have on treatment to improve clinical outcomes. Such studies are essential for implementation of biomarker panels into clinical care. To the best of our knowledge, there is presently only one study assessing the clinical utility of a biomarker panel for this purpose. The PRIORITY trial (NCT 02040441) aims to determine the effect of spironolactone versus placebo in delaying the onset of microalbuminuria in patients with type 2 diabetes at high-risk for diabetic kidney disease progression that are identified by a urinary proteomic biomarker score. Besides the need for clinical studies to assess the impact of using biomarkers to guide treatment

decisions, the reporting of methodological aspects of biomarker studies can be improved. Novel studies should state specific characteristics of the assays that are used and biomarker storage conditions. The impact of prolonged frozen storage of serum and urine on stability of biomarkers, and the effect of (repeated) freeze-thaw cycles on biomarker concentration are key issues that should be addressed. Finally, there is heterogeneity in the statistical methods that are used to report the predictive value of biomarker panels. Though it becomes more common to report novel metrics of model discrimination, such as AUROC and NRI or IDI, different measures were reported in the studies that were reviewed, and harmonization is required. In this respect use of the Net Benefit score might be promising.

Before novel biomarkers or biomarker panels can be used in clinical practice, high quality studies are required that ascertain the accuracy, precision, and predictive ability of the biomarker panel. We have provided key points in design and methodology that should be addressed in such studies. These points may be of help to refine future biomarker panel studies.

References

- 1 Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clin Pharmacol Ther* 69(3): 89-95, 2001.
- 2 KDOQI clinical practice guidelines and clinical practice recommendations for diabetes and chronic kidney disease. *Am J Kidney Dis* 49(2, Supplement 2): S12-S154, 2007.
- 3 Fox CS, Matsushita K, Woodward M, Bilo HJ, Chalmers J, Heerspink HJ, Lee BJ, Perkins RM, Rossing P, Sairenchi T, Tonelli M, Vassalotti JA, Yamagishi K, Coresh J, de Jong PE, Wen CP, Nelson RG, Chronic Kidney Disease Prognosis Consortium. Associations of kidney disease measures with mortality and end-stage renal disease in individuals with and without diabetes: A meta-analysis. *Lancet* 380(9854): 1662-1673, 2012.
- 4 Parving HH, Persson F, Rossing P. Microalbuminuria: A parameter that has changed diabetes care. *Diabetes Res Clin Pract* 2014.
- 5 Pena MJ, de Zeeuw D, Mischak H, Jankowski J, Oberbauer R, Woloszczuk W, Benner J, Dahlman G, Mayer B, Mayer G, Rossing P, Lambers Heerspink HJ. Prognostic clinical and molecular biomarkers of renal disease in type 2 diabetes. *Nephrol Dial Transplant Supp* in press 2015.
- 6 Pena MJ, Heinzel A, Heinze G, Alkhalaf A, Bakker SJL, Nguyen TQ, Goldschmeding R, Bilo HJG, Perco P, Mayer B, de Zeeuw D, Lambers Heerspink HJ. A novel panel of biomarkers to predict eGFR decline in type 2 diabetes. *PLoS ONE* in press.
- 7 Agarwal R, Duffin KL, Laska DA, Voelker JR, Breyer MD, Mitchell PG. A prospective study of multiple protein biomarkers to predict progression in diabetic chronic kidney disease. *Nephrol Dial Transplant* 2014.
- 8 Verhave JC, Bouchard J, Goupil R, Pichette V, Brachemi S, Madore F, Troyanov S. Clinical value of inflammatory urinary biomarkers in overt diabetic nephropathy: A prospective study. *Diabetes Res Clin Pract* 101(3): 333-340, 2013.
- 9 Wong MG, Perkovic V, Woodward M, Chalmers J, Li Q, Hillis GS, Yaghubian Azari D, Jun M, Poulter N, Hamet P, Williams B, Neal B, Mancia G, Cooper M, Pollock CA. Circulating bone morphogenetic protein-7 and transforming growth factor-beta1 are better predictors of renal end points in patients with type 2 diabetes mellitus. *Kidney Int* 83(2): 278-284, 2013.
- 10 Schlatzer D, Maahs DM, Chance MR, Dazard JE, Li X, Hazlett F, Rewers M, Snell-Bergeon JK. Novel urinary protein biomarkers predicting the development of microalbuminuria and renal function decline in type 1 diabetes. *Diabetes Care* 35(3): 549-555, 2012.
- 11 Desai AS, Toto R, Jarolim P, Uno H, Eckardt KU, Kewalramani R, Levey AS, Lewis EF, McMurray JJ, Parving HH, Solomon SD, Pfeffer MA. Association between cardiac biomarkers and the development of ESRD in patients with type 2 diabetes mellitus, anemia, and CKD. *Am J Kidney Dis* 58(5): 717-728, 2011
- 12 Kern EF, Erhard P, Sun W, Genuth S, Weiss MF. Early urinary markers of diabetic kidney disease: A nested case-control study from the diabetes control and complications trial (DCCT). *Am J Kidney Dis* 55(5): 824-834, 2010.
- 13 Astrup AS, Tarnow L, Pietraszek L, Schalkwijk CG, Stehouwer CD, Parving HH, Rossing P. Markers of endothelial dysfunction and inflammation in type 1 diabetic patients with or without diabetic nephropathy followed for 10 years: Association with mortality and decline of glomerular filtration rate. *Diabetes Care* 31(6): 1170-1176, 2008.
- 14 Persson F, Rossing P, Hovind P, Stehouwer CD, Schalkwijk CG, Tarnow L, Parving HH. Endothelial dysfunction and inflammation predict development of diabetic nephropathy in the irbesartan in patients with type 2 diabetes and microalbuminuria (IRMA 2) study. *Scand J Clin Lab Invest* 68(8): 731-738, 2008
- 15 Hlatky MA, Greenland P, Arnett DK, Ballantyne CM, Criqui MH, Elkind MS, Go AS, Harrell FE, Jr, Hong Y, Howard BV, Howard VJ, Hsue PY, Kramer CM, McConnell JP, Normand SL, O'Donnell CJ, Smith SC, Jr, Wilson PW, American Heart Association Expert Panel on Subclinical Atherosclerotic Diseases and Emerging Risk Factors and the Stroke Council. Criteria for evaluation of novel markers of cardiovascular risk: A scientific statement from the American Heart Association. *Circulation* 119(17): 2408-2416, 2009.
- 16 Systems Biology towards Novel Chronic Kidney Disease Diagnosis and Treatment. [Internet] Cited 2014 Dec 31] Available from: <https://www.syskid.eu>
- 17 Levey AS, Inker LA, Matsushita K, Greene T, Willis K, Lewis E, de Zeeuw D, Cheung AK, Coresh J. GFR decline as an end point for clinical trials in CKD: A scientific workshop sponsored by the National Kidney Foundation and the US Food and Drug Administration. *Am J Kidney Dis* 64(6): 821-835, 2014.

- 18 Lambers Heerspink HJ, Kropelin TF, Hoekman J, de Zeeuw D, on behalf of the, Reducing Albuminuria as Surrogate Endpoint (REASSURE) Consortium. Drug-induced reduction in albuminuria is associated with subsequent renoprotection: A meta-analysis. *J Am Soc Nephrol* 2014.
- 19 Prassas I, Brinc D, Farkona S, Leung F, Dimitromanolakis A, Chrystoja CC, Brand R, Kulasingam V, Blasutig IM, Diamandis EP. False biomarker discovery due to reactivity of a commercial ELISA for CUZD1 with cancer antigen CA125. *Clin Chem* 60(2): 381-388, 2014.
- 20 Nauta FL, Bakker SJ, Lambers Heerspink H, de Zeeuw D, van Oeveren W, Bilo H, de Jong PE, Gansevoort RT. Effect of frozen storage on urinary concentration of kidney damage markers. *Am J Kidney Dis* 59(4): 586-589, 2012.
- 21 Brinkman JW, de Zeeuw D, Duker JJ, Gansevoort RT, Kema IP, Hillege HL, de Jong PE, Bakker SJ. Falsely low urinary albumin concentrations after prolonged frozen storage of urine samples. *Clin Chem* 51(11): 2181-2183, 2005.
- 22 Harrel FE: Regression modeling strategies - with applications to linear models, logistic regression, and survival analysis. <http://www.springer.com/mathematics/probability/book/978-0-387-95232-1> ed., New York, Springer, 2001.
- 23 Sauerbrei W. The use of resampling methods to simplify regression models in medical statistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 48(3): 313-329, 1999.
- 24 Hellemons ME, Kerschbaum J, Bakker SJ, Neuwirt H, Mayer B, Mayer G, de Zeeuw D, Lambers Heerspink HJ, Rudnicki M. Validity of biomarkers predicting onset or progression of nephropathy in patients with type 2 diabetes: A systematic review. *Diabet Med* 29(5): 567-577, 2012.
- 25 Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol* 159(9): 882-890, 2004.
- 26 Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 115(7): 928-935, 2007.
- 27 Pencina MJ, D'Agostino RB, Pencina KM, Janssens AC, Greenland P. Interpreting incremental value of markers added to risk prediction models. *Am J Epidemiol* 176(6): 473-481, 2012.
- 28 Leening MJ, Vedder MM, Witteman JC, Pencina MJ, Steyerberg EW. Net reclassification improvement: Computation, interpretation, and controversies: A literature review and clinician's guide. *Ann Intern Med* 160(2): 122-131, 2014.
- 29 Pencina MJ, D'Agostino RB S, D'Agostino RB, Jr, Vasan RS. Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Stat Med* 27(2): 157-72; discussion 207-12, 2008.
- 30 Steyerberg EW, & Vergouwe Y. Towards better clinical prediction models: Seven steps for development and an ABCD for validation. *Eur Heart J* 35(29): 1925-1931, 2014.
- 31 Kerr KF, Meisner A, Thiessen-Philbrook H, Coca SG, Parikh CR. Developing risk prediction models for kidney injury and assessing incremental value for novel biomarkers. *Clin J Am Soc Nephrol* 9(8): 1488-1496, 2014.

