From patient's perspective to predictive modeling: cerebrovascular events and chronic kidney disease

Jong, Y. de

FROM PATIENT'S PERSPECTIVE TO PREDICTIVE MODELING

Cerebrovascular events and chronic kidney disease

Ype de Jong

FROM PATIENT'S PERSPECTIVE TO PREDICTIVE MODELING

Cerebrovascular events and chronic kidney disease

Ype de Jong

The work described in this thesis was performed at the Department of Clinical Epidemiology, Leiden University Medical Center, The Netherlands.

Y. de Jong was supported by a grant from The Dutch Kidney Foundation (Nierstichting, 16OKG12).

Cover design and layout:	Ridderprint | www.ridderprint.nl
Printed by:	Ridderprint | www.ridderprint.nl
ISBN:	978-94-6506-372-0

FROM PATIENT'S PERSPECTIVE TO PREDICTIVE MODELING

Cerebrovascular events and chronic kidney disease

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof. dr. ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op 10 oktober 2024
klokke 11:30 uur

door

Ype de Jong
geboren te Amersfoort
in 1990

Promotores

Prof. Dr. F.W. Dekker

Co-promotor

Dr. M. van Diepen

Leden promotiecommissie

Prof. Dr. E. W. Steyerberg

Prof. Dr. F. A. Klok

Dr. E. K. Hoogeveen

Dr. G. Ocak

The printing fees for this thesis were partly covered by funding from the following organizations. Leiden University Medical Center provided financial support. Financial support of the Dutch Heart Foundation for the publication of this thesis is gratefully acknowledged. Chipsoft kindly provided financial support.

TABLE OF CONTENTS

Chapter 1

GENERAL INTRODUCTION

1.1 INTRODUCTION

"The notion that it will be more difficult in the future is always there. I may not have many problems right now, but the sword of Damocles is always hanging over my head."

This quote from one of the patients with chronic kidney disease (CKD), extracted from a qualitative study discussed in this thesis, illustrates the looming fear of the unknown for patients living with impaired kidney function. Healthy kidneys are instrumental for maintaining homeostasis in many processes: they regulate electrolyte concentration and acid-base balance, excrete toxins and waste products in the urine, play a role in bone mineralization, and manage the blood pressure and extracellular fluid volume. They achieve this both independently and in concert with other – mainly endocrine – organs. Damage to these vital organs, either acute or chronic, consequently result in disturbance of these processes and consequently in a wide diversity of possible outcomes. The chronic aspect of kidney disease, meaning that once present, this damage is persistent and cannot be cured by medication, is troubling for most people who suddenly – at least in the view of healthcare professionals – become *patients*. Having been diagnosed with CKD, as the person quoted above, most patients realize that many possible outcomes may happen to them. The uncertainty *which* outcomes, *when* they might occur and in *what* severity can be paralyzing. These questions, though difficult to answer, are important aspects of chronic care that need to be addressed by the healthcare team involved in the care for these patients.

1.2 LIVING WITH CKD

1.2.1 Progression of CKD: diagnosis, staging and disease trajectories

CKD is a heterogenous group of kidney diseases, sharing risk factors with diseases including, but not limited to diabetes mellitus, hypertension, cardiovascular diseases, and auto-immune disorders. It affects 9.1% (95% CI 8.5-9.8) of the population worldwide, and is increasing in prevalence[1]. Especially in earlier stages, kidney function impairment may be entirely asymptomatic. It may be diagnosed by chance in blood works with a reduced glomerular filtration rate (GFR), or an increased protein excretion in the urine screen. In daily clinical practice, GFR is usually estimated (*estimated* GFR, i.e. eGFR) by the serum creatinine concentration. This can be calculated using formulae that incorporate patients characteristics such as age, sex, and race (for the Modification of Diet in Renal Disease [MDRD][2] and the more commonly used Chronic Kidney Disease Epidemiology Collaboration [CKD-EPI][3]; an updated equation that does not include race as a variable was published recently[4]) and weight (as for the now lesser used Cockcroft-

Gault equation[5]). A 24-hour urinary creatin excretion is often used as confirmatory measurement when decreased GFR is suspected. Alternatively, GFR can be measured (mGFR) using exogenous markers such as inulin, iohexol or radioactive traces – each method with their own strengths and limitations[6]. Some kidney diseases, such as IgA nephropathy or membranous glomerulonephritis may present with urinary protein loss but retained GFR. Thus, measurement of urinary protein in addition to eGFR, either in a random sample or a 24h, is essential in the evaluation of kidney function impairment.

Depending on underlying cause, kidney damage may either resolve (acute kidney injury, AKI) or become chronic (CKD, defined as abnormalities of kidney structure or function, present for ≥3 months, with implications for health[7]). International guidelines, such as the Kidney Disease: Improving Global Outcomes (KDIGO)[7] recommend staging CKD with both the eGFR and urinary protein excretion (**Figure 1.1**) correlating well to CKD related complications. CKD may stabilize, or progress towards end stage kidney disease (ESKD). In this final phase, the patient is counseled on renal replacement therapy (RRT), i.e. peritoneal- or hemodialysis, and the options of receiving a kidney transplantation. Alternatively, conservative (non-dialytic) care management may be offered, which instead of prolonging life, focuses on the quality of life and symptom control[8].

				Persistent albuminuria categories		
				A1	A2	A3
				Normal to mildly increased	Moderately increased	Severely increased
				<30 mg/g	30-300 mg/g	>300 mg/g
GFR categories eGFR range (ml/min/1.73 M2	G1	Normal or high	>90			
	G2	Mildly decreased	60-89			
	G3a	Mildly to moderately decreased	45-59			
	G3b	Moderately to severely decreased	30-44			
	G4	Severely decreased	15-29			
	G5	Kidney failure	<15			

Figure 1.1 CKD categories based on eGFR and albuminuria Data from the Kidney Disease: Improving Global Outcomes (KDIGO) 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease[7]. Colors indicate the risk of various outcomes, such as ESKD, progressive CKD, all-cause and cardiovascular mortality (green: low risk, yellow: moderately increased risk, orange: high risk, red: very high risk). KDIGO advices to stage kidney damage by combining the eGFR and degree of albuminuria (e.g. G4A2, meaning an eGFR of 15-29 in combination with 30-300 mg/g albuminuria).

1.2.2 Risk of ischemic stroke in later stages of CKD

In contrast to the largely asymptomatic nature of early-stage CKD, numerous symptoms related to kidney function impairment may arise with progression of disease, often peaking in ESKD[9]. These symptoms can be categorized as specific to the underlying disease causing CKD (e.g. slow healing wounds may be present in the uncontrolled diabetic patient with CKD), whereas other symptoms such as itching can be caused by mechanisms related to the impaired kidney function itself. Impaired kidney function may also aggravate disease specific risks for certain outcomes. For example, and relevant to this thesis, patients with atherosclerosis and CKD are at increased risk for cardiovascular events. This risk is especially increased for ischemic stroke (IS), with relative risk ratios of 5-30, depending on the demographic factors and kidney function[10-12]. Again, part of the mechanism are shared traditional risk factors for IS and CKD: e.g. advanced age, hypertension, smoking, and diabetes, amongst other risk factors. However, there is a positive correlation between CKD severity and the incidence of IS[10-12], suggesting CKD-specific pathophysiological mechanisms attributing to this risk as well[13]. This pathway from CKD leading to IS is complex, with many different proposed factors accounting for this increased risk[14]. Chronic inflammation, oxidative stress, uremic toxins, thrombogenic factors, abnormalities in the calcium and phosphate homeostasis have been correlated with IS in CKD patients[14]. Prolonged exposure to this may lead to media calcification vessel stiffening, showing a peak in IS incidence in patients reaching ESKD[15]. For patients on dialysis, IS incidence is even further increased, especially shortly after hemodialysis initiation[13 14 16]; for peritoneal dialysis, risk is moderately increased[11].

The other pathway involves cardiac arrhythmias subsequently leading to IS. Patients with CKD are at increased risk for heart rhythm disorders, especially atrial fibrillation (AF). AF is highly prevalent amongst CKD with an estimated prevalence of 12-27% depending on the severity of CKD and dialysis status[13]. Shared risk factors to develop AF in CKD patients include advanced age, ischemic cardiomyopathy, diabetes, amongst others. Risk factors specific to CKD include oxidative stress, chronic volume overload and activation of the renin-aldosterone-angiotensin system. In dialysis patients, these volumetric and electrolyte changes are more acute, precipitating to sympathetic activation and catecholamine release[17-20]. Interestingly, though the mechanism is largely unknown, AF can also cause or further worsen CKD[17]. The pathophysiological relations between kidney function impairment, IS and AF – central for part II of this thesis – are illustrated in **Figure 1.2**.

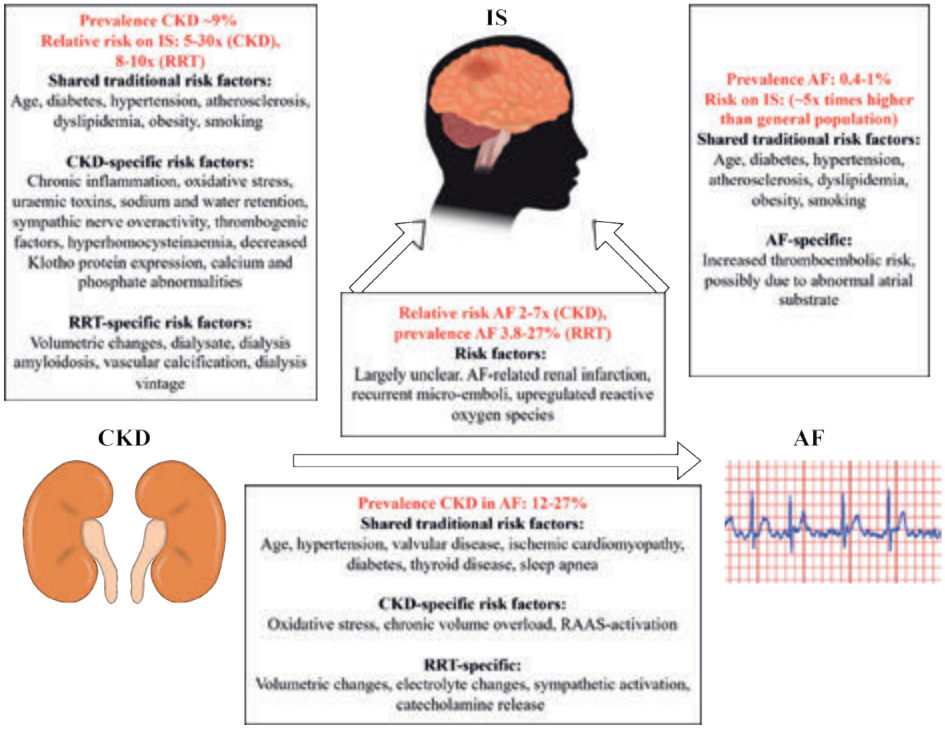


Figure 1.2 Schematic overview of the relation between CKD, IS, and AF, highlighting some proposed pathophysiological pathways. More detail on the specific risk factors of IS in CKD populations can be found in a review by Toyoda *et al*[14] (incidence of IS in CKD is 'dose dependent'; i.e. dependent on the severity of kidney function loss[10-12] and RRT[10-12]); the relation between AF and CKD in the studies by Chen *et al*[17] and Voroneanu *et al*[18] (prevalence of AF in CKD[19] is again dose dependent; and RRT[20]). Details on the relation between AF and IS, and the hypothesis on atrial remodeling and abnormal atrial substrate, which falls beyond the aims of this thesis, can be found in the study by Kamel *et al*[21]. Abbreviations: CKD: chronic kidney disease; AF: atrial fibrillation; IS: ischemic stroke; RAAS: renin-aldosterone-angiotensin system; RRT: renal replacement therapy.

1.2.3 Disease awareness and quality of life

"Having CKD is just like walking in the valley of the shadow of death, and I can see no hope … My children are still so young. Death has cast a shadow over me, and I am very affected."

Symptom burden, and the effect on health related quality of life (HRQOL) of patients on dialysis are comparable to patients with other impactful chronic diseases, such as cancer[22]. However, the absence of symptoms in the early stages of CKD, in combination with the low awareness of both normal kidney function and kidney disease in the general population[22-24], makes it difficult for patients with CKD to cope and take control of their disease[25-27]. As the underlying disease progresses and kidney function deteriorates, symptoms may arise.

In this phase, triggered by symptoms they had not experienced before, many patients realize the consequence of living with a chronic disease[28].

In an attempt to influence the trajectory of their disease, patients express the urge to gain control on their disease, and regain control on their life. Consequently, in this phase, healthcare visits are usually more frequent[29], patients seek information on both the disease and the possible treatments, and prioritize those aspects of care that are important to them[30]. For example, while for younger patients reaching kidney failure and RRT is a pivotal moment[31], for elderly patients quality of life may be more important than duration of life[30]. Other aspects of CKD include experienced health and symptom burden, the effects of a chronic disease on social life and the effects of disease on economic productivity. Yet, it is increasingly acknowledged that these aspects, often prioritized by the patient, are not always adequately reflected by the priorities of their healthcare provider[31 32]. Patients' beliefs, knowledge and feelings play a crucial part in the consultation with medical professionals. Understanding the patient's reasons for consultation (which might be different from the medical professional's), and also the expectations of the patient, requires the healthcare provider to 'see the disease through the patient's eye'[33 34]. Failure to do so has been shown to negatively influence patient satisfaction, drug compliance, and the feeling of being in control on chronic diseases[33]. Traditionally, patient centeredness has received little attention in the medical curriculum, which is often focused on the medical aspects such prolongation of time to kidney failure or death, and less on HRQOL or palliative care[33]. For example, in a survey among nephrology fellows, nearly all acknowledged the importance of HRQOL and palliative care, but felt unready to discuss these topics properly because of lack of training[35]. In a follow-up survey ten years later (among different residents), the same urgency for education on the management of HRQOL and palliative care was experienced, but the perception of the quality thereof remained similarly poor[36]. Consequently, patients with ESKD appear to receive lower quality end-of-life care than patients with other chronic diseases, such as cancer, cardiovascular diseases, or dementia[37 38].

1.2.4 Understanding why: qualitative studies

Besides incongruency between the aspects in care prioritized by patients and healthcare professionals, underreporting of symptoms by patients, or not recognizing symptoms by the healthcare professional is another reason for suboptimal symptom management. This may stem for example from an earlier experience of feeling unheard, or the professional's focus on pathophysiological parameters instead of symptoms[39]. Without knowledge of which and why certain aspects of their disease matter to patients, implementation of

patient centered care makes little sense[39]. Qualitative studies are well fit to provide context for such questions. By obtaining data from e.g. interviews, observations or focus groups, the qualitative researcher seeks to determine the patient's view of certain aspects in healthcare by means of description. It can be regarded as hypothesis generating research, and though fundamentally different from a methodological standpoint when compared to quantitative research, it is not necessarily incompatible nor mutually exclusive: both methods can complement each other. For example, findings from a quantitative study can be explained by qualitative research, or hypotheses generated in a qualitative study can be statistically tested in quantitative research. The value of these studies are increasingly recognized in healthcare[40], for example to explore the difference in priorities between patients and doctors[41], or decision-making processes for patients with ESKD in regard to choosing between RRT and conservative care. In addition, such qualitative studies provide a rich context for the development of more quantifiable outcomes, such as patient reported outcomes measures (PROM).

Reviews of qualitative studies can result in a greater conceptual understanding of the topic beyond the single studies[42], and across different contexts[43]. They can thus serve as a method to generate new theoretical or conceptual models, or more pragmatically, provide information for the implementation of e.g. health interventions[43]. Compared to qualitative studies, this method of data aggregation is relatively novel, with only a handful publications of this type prior the 2000s, followed by an exponential growth in the two decades thereafter. In this evolving field, guidance on reporting was quickly provided by the introduction of the ENTREQ (Enhancing transparency in reporting the synthesis of qualitative research) guideline in 2012[43]. It replaced the checklist on single qualitative studies, the COREQ (Consolidated Criteria for Reporting Qualitative Research)[44], which was sometimes used for that purpose in the absence of a dedicated checklist.

1.3 PREDICTION RESEARCH IN CKD

"I've got a rough timeframe, again it's imperfect so no one knows definitively. People say 'When are you going on dialysis?' Well no-one knows but we can guess the way it's going, we can guess"[45]

Living with the diagnosis CKD is marked by uncertainty about what to expect. Uncertainty on the prognosis, caused for example by unclarity on the etiology or the effectiveness of therapy, often results in vagueness on this topic in the outpatient clinic. In the context of these uncertainties, healthcare professionals tend to avoid discussions on prognosis, fearing for example that it might overwhelm the patient or hamper the patient-doctor relation[35]

[46 47]. Yet, patients have a realistic expectation that neither the risk of future outcomes nor the timeframe of reaching them can be given with certainty and urge the need to present this information nonetheless, as it helps to regain control[48 49].

Prediction research plays a key role in providing prognostic information tailored to the specific characteristics of the patient at hand[50]. In that sense, prediction models can play an important role in patient-centered care provision. Using a prediction model – a mathematical equation, developed on patient data – the clinician can calculate the risk of an outcome, and for example discuss this risk in the context of choosing between treatment modalities for ESKD. Within the field of CKD and ESKD, the focus of prediction model studies has traditionally been on the risk of ESKD in CKD patients[51], and death in ESKD patients[52]. Unfortunately, as in many other medical specialties, implementation of prediction models in routine care is hampered by methodological issues, risk of bias and incomplete reporting[53]. Recent years have seen a massive increase in the number of models developed, however, external validation of these models – an essential benchmarking test – is often not performed[53 54]. External validation is the process of testing the predictive capabilities of the developed model in a different population than the development cohort to test its transportability. Two measures of predictive performance are essential to predictive research: discrimination, which describes the model's ability to assign higher risk to those reaching the outcome than to those who do not, and calibration, which assesses the agreement between the predicted risks and observed frequencies[55]. Simply obtaining this information by validating the model in the development data is insufficient: as the model has been precisely fitted to this data, it may show excellent predictive capabilities, a statistical phenomenon known as optimism. Validating a model in the development data – known as internal validation – can provide insight in how the model will perform in an external (i.e. not in the development cohort) target population. Different methods are available, such as split sample (the model is developed in part of the population, and validated in the other), cross-validation (developing the model in a random subset, and validating it in the remaining; repeating this procedure multiple times), and bootstrapping (replicating the sampling procedure, by drawing samples with replacement; repeating this a large number of times)[56]. Generally speaking however, without an external validation, it is impossible to estimate the accuracy of the predicted probabilities in the target population[29].

1.3.1 Stroke risk prediction in CKD.

Besides these issues, it remains an unanswered question whether the traditionally predicted outcomes – ESKD and death – are regarded equally important by the clinician

as the patient. Patient centered care involves discussing the possible outcomes of CKD – and there are many possible outcomes besides ESKD and death. For example, as discussed above, cardiovascular events such as IS constitute to high risk of morbidity. Yet, besides the 42 models for ESKD and 16 models for predicting mortality we identified previously[51 52], models for other outcomes prioritized by patients with CKD or ESKD are scarcely developed. No models for IS have been developed in CKD or ESKD populations. Consequently, the current guideline-endorsed models predicting IS such as the CHA_2DS_2-VASc (Congestive heart failure or left ventricular dysfunction Hypertension, Age ≥75, Diabetes, Stroke, Vascular disease, Age 65–74, Sex category)[60] are being used by clinicians in daily clinical practice. Especially in the setting where IS risk is weighed with the risk of therapy-related bleeding, correct risk estimations are crucial. In the absence of data on the external validity of these models in the high risk CKD and ESKD populations, the risk of misclassification is high: overprediction will result in overtreatment and consequently higher bleeding rates; while underprediction will result in higher stroke rates[61].

1.3.2 Validation studies: risk of bias

Validating a model in a different target population than the source population, may affect the predictive performance via various mechanisms. For example, a model developed in a heterogeneous population – say IS risk in AF patients – is likely to perform differently in a more homogeneous population, such as AF patients on hemodialysis. This difference in patients characteristics, or case-mix, negatively affects discrimination[62]. However, this reduced predictive performance is still the 'true', or unbiased performance of the model in the target population. That is, assuming that the validation was correctly performed, which is not the case for most prediction studies. Commonly encountered methodological flaws that may result in biased outcomes include differences in predictor- or outcome definitions, differences in prediction windows (i.e. the time between prediction and the moment the outcome is assessed), inadequate sample size, or incorrect handling of missing data.

While risk of bias (ROB) in the setting of etiologic research could be defined relatively intuitively as a systematic error that may affect the study's validity, little evidence regarding bias in prediction research exists. One way to look at bias in this context is to assess the differences in predictive performance in the development versus the validation study: apert differences might indicate ROB in the developed model. As illustrated however, there may be legitimate reasons for the drop in predictive performance. The authors of a recently introduced guideline for ROB in prediction research, the PROBAST (Prediction model Risk Of Bias ASsessment Tool) define bias as "(…) shortcomings in study design, conduct, or analysis that could lead to systematically distorted estimates of

a model's predictive performance"[63 64]. This, still, is a difficult definition: without a gold standard to compare the predictive performance, how should one know one's estimates are distorted? Though an overviewing article of ROB in prediction research and the trends over time is currently lacking, key reviews in different fields of clinical medicine suggest a high prevalence of bias in prediction studies[51 52 61 64-70]. This notion, despite the enormous potential of predictive modelling, warrants caution for the use of this type of research in clinical medicine – and especially so for unvalidated models.

1.4 AIMS AND OUTLINE

The title of this thesis summarizes the central aim: to explore the perspectives of the individual patient with CKD regarding outcomes that are important to him, and subsequently predict such outcomes by means of a prediction model. This aim is undoubtedly ambitious—perhaps overly so for a single thesis—but, as we will argue later, it could be the crucial method for implementing person-centered care. This thesis is structured in two parts. In **part I**, using qualitative methods, we explore the views of patients living with CKD on their prognosis as part of patient centered care implementation. In **part II**, we study the predictive performance of prediction models for IS in patients with CKD and ESKD. In both parts, we discuss methodological considerations of these different study types, with a focus on quality of reporting and ROB.

PART I: Patient's perspective on their prognosis & methodological considerations of qualitative research

Patients diagnosed with CKD face the challenge of coping with their disease, and quickly discover that necessary information – both regarding the etiology, treatment and future outcomes – is not tailored to their needs. In **Chapter 2**, using a qualitative systematic review design, we explore the views of patients with CKD on their prognosis, and explore which outcomes they prioritize. Furthermore, we look into barriers and facilitators for the implementation of patient centered care in CKD.

Historically, the relative novelty of this study methodology (i.e. qualitative reviews) resulted in unstandardized reporting. Though the introduction of the COREQ in 2007 for qualitative studies was certainly helpful for qualitative reviews, the publication of the ENTREQ guidelines in 2012 aimed to provide the definitive guidance for reporting. In **Chapter 3**, we explore the uptake of both guidelines in qualitative reviews in all clinical fields since the first publication of this study type. Next, we study the impact of the COREQ on the quality of reporting of single qualitative studies.

PART II: Prediction of ischemic stroke in CKD and dialysis: methodological quality of current models and clinical implications

Prediction models for patients with CKD or those receiving RRT focus mainly on reaching ESKD or death, respectively. In the first part, we focused on prognostic uncertainty of these patients regarding various outcomes. In this part, we focus on the prediction of ischemic stroke in CKD and RRT – one of the major risks of CKD. While it is certainly tempting to use predictive models in this setting, limited information on the predictive performance of models such as the CHA_2DS_2-VASc is available. The current use of these models comes with the risk of misclassification: overpredicting the outcome results in overtreatment and subsequent increased incidence of bleeding, and vise-versa for underprediction.

In **Chapter 4**, to identify which prediction models have been published so far to predict IS, we conduct a systematic review. Next, we assess and compare the predictive performance of these models in a large incident dialysis population. In this external validation we also look into the methodological quality of these models, assessing the risk of bias using the PROBAST.

In **Chapter 5**, we again validate multiple models for IS, but this time in a large group Swedish patients with incident AF, with and without CKD. The aim of this chapter is to identify the most stable model across the whole spectrum of kidney function. As discussed above, in the section on risk of bias, models are often validated on a different prediction window than they have been developed. We look into the stability of both discrimination an calibration in the large when this prediction window is not acknowledged, by prolonging the prediction window.

Finally, in **Chapter 6**, we discuss ROB in the context of prediction research in all clinical fields, using, but also extending on the PROBAST. Illustrating the prevalence of bias in prediction research, and trends over time, we conduct a scoping review on reviews that used the PROBAST for ROB assessment.

1.4 REFERENCES

1. Global, regional, and national burden of chronic kidney disease, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2020;395(10225):709-33. doi: 10.1016/s0140-6736(20)30045-3 [published Online First: 2020/02/18]
2. Levey AS, Coresh J, Greene T, et al. Using standardized serum creatinine values in the modification of diet in renal disease study equation for estimating glomerular filtration rate. *Ann Intern Med* 2006;145(4):247-54. doi: 10.7326/0003-4819-145-4-200608150-00004
3. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med* 2009;150(9):604-12. doi: 10.7326/0003-4819-150-9-200905050-00006
4. Inker LA, Eneanya ND, Coresh J, et al. New Creatinine- and Cystatin C–Based Equations to Estimate GFR without Race. *New England Journal of Medicine* 2021;385(19):1737-49. doi: 10.1056/NEJMoa2102953
5. Cockcroft DW, Gault MH. Prediction of creatinine clearance from serum creatinine. *Nephron* 1976;16(1):31-41. doi: 10.1159/000180580
6. Levey AS, Coresh J, Tighiouart H, et al. Measured and estimated glomerular filtration rate: current status and future directions. *Nature Reviews Nephrology* 2020;16(1):51-64. doi: 10.1038/s41581-019-0191-y
7. Stevens PE, Levin A. Evaluation and management of chronic kidney disease: synopsis of the kidney disease: improving global outcomes 2012 clinical practice guideline. *Ann Intern Med* 2013;158(11):825-30. doi: 10.7326/0003-4819-158-11-201306040-00007
8. Murtagh FE, Burns A, Moranne O, et al. Supportive Care: Comprehensive Conservative Care in End-Stage Kidney Disease. *Clin J Am Soc Nephrol* 2016;11(10):1909-14. doi: 10.2215/cjn.04840516 [published Online First: 2016/08/12]
9. de Rooij ENM, Meuleman Y, de Fijter JW, et al. Symptom Burden before and after Dialysis Initiation in Older Patients. *Clin J Am Soc Nephrol* 2022;17(12):1719-29. doi: 10.2215/cjn.09190822 [published Online First: 2022/11/11]
10. Kelly DM, Rothwell PM. Does Chronic Kidney Disease Predict Stroke Risk Independent of Blood Pressure?: A Systematic Review and Meta-Regression. *Stroke* 2019;50(11):3085-92. doi: 10.1161/strokeaha.119.025442 [published Online First: 2019/10/09]
11. Nayak-Rao S, Shenoy MP. Stroke in Patients with Chronic Kidney Disease…: How do we Approach and Manage it? *Indian J Nephrol* 2017;27(3):167-71. doi: 10.4103/0971-4065.202405 [published Online First: 2017/05/30]
12. Lee M, Saver JL, Chang KH, et al. Low glomerular filtration rate and risk of stroke: meta-analysis. *Bmj* 2010;341:c4249. doi: 10.1136/bmj.c4249 [published Online First: 2010/10/05]
13. Findlay MD, Thomson PC, Fulton RL, et al. Risk Factors of Ischemic Stroke and Subsequent Outcome in Patients Receiving Hemodialysis. *Stroke* 2015;46(9):2477-81. doi: 10.1161/strokeaha.115.009095 [published Online First: 2015/08/01]
14. Toyoda K, Ninomiya T. Stroke and cerebrovascular diseases in patients with chronic kidney disease. *Lancet Neurol* 2014;13(8):823-33. doi: 10.1016/s1474-4422(14)70026-2 [published Online First: 2014/07/18]
15. Dube P, DeRiso A, Patel M, et al. Vascular Calcification in Chronic Kidney Disease: Diversity in the Vessel Wall. *Biomedicines* 2021;9(4) doi: 10.3390/biomedicines9040404 [published Online First: 2021/05/01]
16. Zhan X, Yang M, Chen Y, et al. Comparison of risk of stroke in patients treated with peritoneal dialysis and hemodialysis: a systematic review and meta-analysis. *Ren Fail* 2019;41(1):650-56. doi: 10.1080/0886022x.2019.1632210 [published Online First: 2019/07/13]

17. Chen TH, Chu YC, Ou SM, et al. Associations of atrial fibrillation with renal function decline in patients with chronic kidney disease. *Heart* 2022;108(6):438-44. doi: 10.1136/heartjnl-2021-319297 [published Online First: 2021/07/02]

18. Voroneanu L, Ortiz A, Nistor I, et al. Atrial fibrillation in chronic kidney disease. *Eur J Intern Med* 2016;33:3-13. doi: 10.1016/j.ejim.2016.04.007 [published Online First: 2016/05/09]

19. Baber U, Howard VJ, Halperin JL, et al. Association of chronic kidney disease with atrial fibrillation among adults in the United States: REasons for Geographic and Racial Differences in Stroke (REGARDS) Study. *Circ Arrhythm Electrophysiol* 2011;4(1):26-32. doi: 10.1161/circep.110.957100 [published Online First: 2010/11/16]

20. Abuhasira R, Mizrakli Y, Shimony A, et al. Atrial Fibrillation Characteristics in Patients on Haemodialysis vs. Peritoneal Dialysis. *Sci Rep* 2018;8(1):2976. doi: 10.1038/s41598-018-21229-9 [published Online First: 2018/02/16]

21. Kamel H, Okin PM, Elkind MS, et al. Atrial Fibrillation and Mechanisms of Stroke: Time for a New Model. *Stroke* 2016;47(3):895-900. doi: 10.1161/strokeaha.115.012004 [published Online First: 2016/01/21]

22. Weisbord SD. Symptoms and their correlates in chronic kidney disease. *Adv Chronic Kidney Dis* 2007;14(4):319-27. doi: 10.1053/j.ackd.2007.07.004 [published Online First: 2007/10/02]

23. Gheewala PA, Peterson GM, Zaidi STR, et al. Public knowledge of chronic kidney disease evaluated using a validated questionnaire: a cross-sectional study. *BMC Public Health* 2018;18(1):371. doi: 10.1186/s12889-018-5301-4 [published Online First: 2018/03/21]

24. Chow KM, Szeto CC, Kwan B, et al. Public lacks knowledge on chronic kidney disease: telephone survey. *Hong Kong Med J* 2014;20(2):139-44. doi: 10.12809/hkmj134134 [published Online First: 2014/03/15]

25. Tuot DS, Plantinga LC. What patients don't know may hurt them: knowledge and the perception of knowledge among patients with CKD. *Kidney Int* 2011;80(12):1256-7. doi: 10.1038/ki.2011.269

26. Wright JA, Wallston KA, Elasy TA, et al. Development and results of a kidney disease knowledge survey given to patients with CKD. *Am J Kidney Dis* 2011;57(3):387-95. doi: 10.1053/j.ajkd.2010.09.018 [published Online First: 20101218]

27. Finkelstein FO, Story K, Firanek C, et al. Perceived knowledge among patients cared for by nephrologists about chronic kidney disease and end-stage renal disease therapies. *Kidney Int* 2008;74(9):1178-84. doi: 10.1038/ki.2008.376 [published Online First: 20080730]

28. Teasdale EJ, Leydon G, Fraser S, et al. Patients' Experiences After CKD Diagnosis: A Meta-ethnographic Study and Systematic Review. *Am J Kidney Dis* 2017;70(5):656-65. doi: 10.1053/j.ajkd.2017.05.019 [published Online First: 2017/08/03]

29. Khan SS, Kazmi WH, Abichandani R, et al. Health care utilization among patients with chronic kidney disease. *Kidney Int* 2002;62(1):229-36. doi: 10.1046/j.1523-1755.2002.00432.x [published Online First: 2002/06/26]

30. Ramer SJ, McCall NN, Robinson-Cohen C, et al. Health Outcome Priorities of Older Adults with Advanced CKD and Concordance with Their Nephrology Providers' Perceptions. *J Am Soc Nephrol* 2018;29(12):2870-78. doi: 10.1681/asn.2018060657 [published Online First: 2018/11/06]

31. González AM, Gutman T, Lopez-Vargas P, et al. Patient and Caregiver Priorities for Outcomes in CKD: A Multinational Nominal Group Technique Study. *Am J Kidney Dis* 2020;76(5):679-89. doi: 10.1053/j.ajkd.2020.03.022 [published Online First: 2020/06/04]

32. Tong A, Jesudason S, Craig JC, et al. Perspectives on pregnancy in women with chronic kidney disease: systematic review of qualitative studies. *Nephrol Dial Transplant* 2015;30(4):652-61. doi: 10.1093/ndt/gfu378 [published Online First: 2014/12/20]

33. Law SA, Britten N. Factors that influence the patient centredness of a consultation. *Br J Gen Pract* 1995;45(399):520-4.

34. Little P, Everitt H, Williamson I, et al. Preferences of patients for patient centred approach to consultation in primary care: observational study. *BMJ* 2001;322(7284):468. doi: 10.1136/bmj.322.7284.468

35. Holley JL, Carmody SS, Moss AH, et al. The need for end-of-life care training in nephrology: national survey results of nephrology fellows. *Am J Kidney Dis* 2003;42(4):813-20. doi: 10.1016/s0272-6386(03)00868-0

36. Combs SA, Culp S, Matlock DD, et al. Update on end-of-life care training during nephrology fellowship: a cross-sectional national survey of fellows. *Am J Kidney Dis* 2015;65(2):233-9. doi: 10.1053/j.ajkd.2014.07.018 [published Online First: 20140920]

37. Wachterman MW, Pilver C, Smith D, et al. Quality of End-of-Life Care Provided to Patients With Different Serious Illnesses. *JAMA Intern Med* 2016;176(8):1095-102. doi: 10.1001/jamainternmed.2016.1200 [published Online First: 2016/07/02]

38. Wong SP, Kreuter W, O'Hare AM. Treatment intensity at the end of life in older adults receiving long-term dialysis. *Arch Intern Med* 2012;172(8):661-3; discussion 63-4. doi: 10.1001/archinternmed.2012.268 [published Online First: 2012/04/25]

39. van der Willik EM, Milders J, Bart JAJ, et al. Discussing results of patient-reported outcome measures (PROMs) between patients and healthcare professionals in routine dialysis care: a qualitative study. *BMJ Open* 2022;12(11):e067044. doi: 10.1136/bmjopen-2022-067044 [published Online First: 2022/11/18]

40. Al-Busaidi ZQ. Qualitative research and its uses in health care. *Sultan Qaboos Univ Med J* 2008;8(1):11-9.

41. Barry CA, Bradley CP, Britten N, et al. Patients' unvoiced agendas in general practice consultations: qualitative study. *Bmj* 2000;320(7244):1246-50. doi: 10.1136/bmj.320.7244.1246

42. Dixon-Woods M, Agarwal S, Jones D, et al. Synthesising qualitative and quantitative evidence: a review of possible methods. *J Health Serv Res Policy* 2005;10(1):45-53. doi: 10.1177/135581960501000110 [published Online First: 2005/01/26]

43. Tong A, Flemming K, McInnes E, et al. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. *BMC Med Res Methodol* 2012;12:181. doi: 10.1186/1471-2288-12-181 [published Online First: 2012/11/29]

44. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care* 2007;19(6):349-57. doi: 10.1093/intqhc/mzm042 [published Online First: 2007/09/18]

45. Campbell-Crofts S, Stewart G. How perceived feelings of "wellness" influence the decision-making of people with predialysis chronic kidney disease. *J Clin Nurs* 2018;27(7-8):1561-71. doi: 10.1111/jocn.14220 [published Online First: 2017/12/15]

46. Heyland DK, Allan DE, Rocker G, et al. Discussing prognosis with patients and their families near the end of life: impact on satisfaction with end-of-life care. *Open Med* 2009;3(2):e101-10. [published Online First: 2009/12/01]

47. Greer RC, Crews DC, Boulware LE. Challenges perceived by primary care providers to educating patients about chronic kidney disease. *J Ren Care* 2012;38(4):174-81. doi: 10.1111/j.1755-6686.2012.00323.x [published Online First: 2012/11/28]

48. Eraker SA, Kirscht JP, Becker MH. Understanding and improving patient compliance. *Ann Intern Med* 1984;100(2):258-68. doi: 10.7326/0003-4819-100-2-258 [published Online First: 1984/02/01]

49. Davison SN, Simpson C. Hope and advance care planning in patients with end stage renal disease: qualitative interview study. *Bmj* 2006;333(7574):886. doi: 10.1136/bmj.38965.626250.55 [published Online First: 2006/09/23]

50. Moons KG, Royston P, Vergouwe Y, et al. Prognosis and prognostic research: what, why, and how? *Bmj* 2009;338:b375. doi: 10.1136/bmj.b375 [published Online First: 2009/02/25]

51. Ramspek CL, de Jong Y, Dekker FW, et al. Towards the best kidney failure prediction tool: a systematic review and selection aid. *Nephrol Dial Transplant* 2020;35(9):1527-38. doi: 10.1093/ndt/gfz018

52. Ramspek CL, Voskamp PW, van Ittersum FJ, et al. Prediction models for the mortality risk in chronic dialysis patients: a systematic review and independent external validation study. *Clin Epidemiol* 2017;9:451-64. doi: 10.2147/clep.S139748 [published Online First: 20170905]

53. Ramspek CL, Jager KJ, Dekker FW, et al. External validation of prognostic models: what, why, how, when and where? *Clin Kidney J* 2021;14(1):49-58. doi: 10.1093/ckj/sfaa188 [published Online First: 2021/02/11]

54. Siontis GC, Tzoulaki I, Castaldi PJ, et al. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J Clin Epidemiol* 2015;68(1):25-34. doi: 10.1016/j.jclinepi.2014.09.007 [published Online First: 2014/12/03]

55. Altman DG, Vergouwe Y, Royston P, et al. Prognosis and prognostic research: validating a prognostic model. *Bmj* 2009;338:b605. doi: 10.1136/bmj.b605 [published Online First: 2009/05/30]

56. Steyerberg EW, Harrell FE, Borsboom GJJM, et al. Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *Journal of Clinical Epidemiology* 2001;54(8):774-81. doi: https://doi.org/10.1016/S0895-4356(01)00341-9

57. Hindricks G, Potpara T, Dagres N, et al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. *Eur Heart J* 2021;42(5):373-498. doi: 10.1093/eurheartj/ehaa612

58. Lip GYH, Banerjee A, Boriani G, et al. Antithrombotic Therapy for Atrial Fibrillation: CHEST Guideline and Expert Panel Report. *Chest* 2018;154(5):1121-201. doi: 10.1016/j.chest.2018.07.040 [published Online First: 2018/08/26]

59. January CT, Wann LS, Calkins H, et al. 2019 AHA/ACC/HRS Focused Update of the 2014 AHA/ACC/HRS Guideline for the Management of Patients With Atrial Fibrillation: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society in Collaboration With the Society of Thoracic Surgeons. *Circulation* 2019;140(2):e125-e51. doi: 10.1161/cir.0000000000000665 [published Online First: 2019/01/29]

60. Lip GY, Nieuwlaat R, Pisters R, et al. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. *Chest* 2010;137(2):263-72. doi: 10.1378/chest.09-1584 [published Online First: 2009/09/19]

61. van der Endt VHW, Milders J, Penning de Vries BBL, et al. Comprehensive comparison of stroke risk score performance: a systematic review and meta-analysis among 6 267 728 patients with atrial fibrillation. *Europace* 2022 doi: 10.1093/europace/euac096 [published Online First: 2022/07/28]

62. Vergouwe Y, Moons KG, Steyerberg EW. External validity of risk models: Use of benchmark values to disentangle a case-mix effect from incorrect coefficients. *Am J Epidemiol* 2010;172(8):971-80. doi: 10.1093/aje/kwq223 [published Online First: 2010/09/03]

63. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Ann Intern Med* 2019;170(1):W1-w33. doi: 10.7326/m18-1377

64. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med* 2019;170(1):51-58. doi: 10.7326/m18-1376

65. Bellou V, Belbasis L, Konstantinidis AK, et al. Prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease: systematic review and critical appraisal. *Bmj* 2019;367:l5358. doi: 10.1136/bmj.l5358 [published Online First: 2019/10/06]

66. Oswald NK, Halle-Smith J, Mehdi R, et al. Predicting Postoperative Lung Function Following Lung Cancer Resection: A Systematic Review and Meta-analysis. *EClinicalMedicine* 2019;15:7-13. doi: 10.1016/j.eclinm.2019.08.015 [published Online First: 2019/11/12]

67. Bracher-Smith M, Crawford K, Escott-Price V. Machine learning for genetic prediction of psychiatric disorders: a systematic review. *Mol Psychiatry* 2021;26(1):70-79. doi: 10.1038/s41380-020-0825-2 [published Online First: 2020/06/28]

68. Perry BI, Upthegrove R, Crawford O, et al. Cardiometabolic risk prediction algorithms for young people with psychosis: a systematic review and exploratory analysis. *Acta Psychiatr Scand* 2020;142(3):215-32. doi: 10.1111/acps.13212 [published Online First: 2020/07/13]

69. Abbasgholizadeh Rahimi S, Légaré F, Sharma G, et al. Application of Artificial Intelligence in Community-Based Primary Health Care: Systematic Scoping Review and Critical Appraisal. *J Med Internet Res* 2021;23(9):e29839. doi: 10.2196/29839 [published Online First: 2021/09/04]

70. Alhmoud B, Bonnici T, Patel R, et al. Performance of universal early warning scores in different patient subgroups and clinical settings: a systematic review. *BMJ Open* 2021;11(4):e045849. doi: 10.1136/bmjopen-2020-045849 [published Online First: 2021/04/08]

1

PART I

The patient's perspective on prognosis
&
Methodological considerations of qualitative research

Chapter 2

PERSON CENTERED CARE PROVISION AND CARE PLANNING IN CHRONIC KIDNEY DISEASE: WHICH OUTCOMES MATTER? A SYSTEMATIC REVIEW AND THEMATIC SYNTHESIS OF QUALITATIVE STUDIES

Ype de Jong
Esmee M. van der Willik
Jet Milders
Yvette Meuleman
Rachael L. Morton
Friedo W. Dekker
Merel van Diepen

BMC Nephrology 2021 Sep 13;22(1):309
DOI: 10.1186/s12882-021-02489-6
PMID: 34517825
PMCID: PMC8438879

To minimize the impact of printing on the environment, the supplementary material has been uploaded. For Chapter 2, the supplement is 19 pages, and can be downloaded using the QR code.

Abstract

Rationale & objective: Explore priorities related to outcomes and barriers of adults with chronic kidney disease (CKD) regarding person centered care and care planning.

Study design: Systematic review of qualitative studies.

Search strategy & sources: In July 2018 six bibliographic databases, and reference lists of included articles were searched for qualitative studies that included adults with CKD stages 1-5, not on dialysis or conservative management, without a previous kidney transplantation.

Analytical approach: Three independent reviewers extracted and inductively coded data using thematic synthesis. Reporting quality was assessed using the COREQ and the review reported according to PRISMA and ENTREQ statements.

Results: Forty-six studies involving 1493 participants were eligible. The period after diagnosis of CKD is characterized by feelings of uncertainty, social isolation, financial burden, resentment and fear of the unknown. Patients show interest in ways to return to normality and remain in control of their health in order to avoid further deterioration of kidney function. However, necessary information is often unavailable or incomprehensible. Although patients and healthcare professionals share the predominant interest of whether or not dialysis or transplantation is necessary, patients value many more outcomes that are often unrecognized by their healthcare professionals. We identified 4 themes with 6 subthemes that summarize these findings: 'pursuing normality and control' ('pursuing normality'; 'a search for knowledge'); 'prioritizing outcomes' ('reaching kidney failure'; 'experienced health'; 'social life'; 'work and economic productivity'); 'predicting the future'; and 'realizing what matters'. Reporting quality was moderate for most included studies.

Limitations: Exclusion of non-English articles.

Conclusions: The realization that patients' priorities do not match those of the healthcare professionals, in combination with the prognostic ambiguity, confirms fatalistic perceptions of not being in control when living with CKD. These insights may contribute to greater understanding of patients' perspectives and a more person-centered approach in healthcare prioritization and care planning within CKD care.

2.1 INTRODUCTION

Chronic kidney disease (CKD) is a group of kidney diseases in which there usually is a gradual decrease in kidney function leading to kidney failure. The often asymptomatic nature of CKD, combined with the low awareness of kidney function in general[1], makes it difficult for patients to comprehend, cope and finally take control after the diagnosis of CKD[2-12]. During the progression of CKD to kidney failure however, numerous physical and psychosocial symptoms may develop, overall reducing health-related quality of life (HRQOL)[13,14]. In this phase, kidney replacement therapy (KRT; kidney transplantation or dialysis), or alternatively conservative therapy is necessary, requiring an informed decision with knowledge of the disease, the possible outcomes and the chances of reaching these outcomes in combination with prioritization of what matters to the patient.

However, for most patients the period between CKD and kidney failure is marked by confusion about the disruptive transition from their pre- to their postdiagnosis self, and uncertainty about what to expect[10,15,16]. Furthermore, it is increasingly acknowledged that outcomes prioritized by clinicians, such as planning for dialysis or transplant, and postponing kidney failure and death, do not adequately reflect patients' desired outcomes, which in contrast may include patient reported outcomes (PROs) like HRQOL or symptom burden[9,17,18]. PRO measures (PROMs) have been developed to further implement person-centered care, by providing insight into outcomes and enhancing the patient-professional conversation about patients' needs and expectations. Such aspects of person-centered care show promising results but have yet to be fully incorporated into routine nephrological care[9, 19-21].

In-depth knowledge about what matters to patients can also be obtained through qualitative research. Moreover, by using qualitative methods, answers to why patients value these outcomes can also be obtained, hereby providing an opportunity for deeper understanding of their motivations, behavior and beliefs. Though frequently used as a first step for the development of PROMs, transferability to other populations than the study subjects of single qualitative studies remains a concern. Systematically reviewing and thematically synthesizing the data of these single studies can result in a greater conceptual understanding of the topic beyond the single studies[22].

Although person-centered care within CKD shows promising results[23-25], better understanding of patients' perspectives on what is important in nephrological care may help to further implement person-centered care. Hence, the aim of this study is to identify outcomes prioritized by patients with CKD, and barriers to person-centered care and care planning, by means of a systematic review and inductive thematic synthesis of qualitative studies among patients with CKD.

2.2 METHOD

We followed the ENTREQ (Enhancing transparency in reporting the synthesis of qualitative research)[26] checklist and the PRISMA[27] statements for reporting our qualitative thematic synthesis.

2.2.1 Selection criteria

All types of written qualitative studies in patients with CKD were included, where data had been collected via interviews, focus groups, or observations. Non-English articles were excluded to prevent cultural and linguistic bias in translation: the original meaning and interpretation may be lost in translation[28]. No publication date constraints were applied. We aimed to identify the priorities regarding outcomes and processes of care and barriers regarding person centered care of patients with CKD, not receiving KRT or conservative management and without a previous kidney transplant. Therefore, studies with mixed populations or mixed methods were excluded if the qualitative data related to patients with CKD was not presented separately. We excluded studies on children (<18 years of age) because of different implications in shared decision making.

2.2.2 Search methods and study selection

Systematically searching for qualitative studies aiming to identify all available studies, instead of a representative sample is problematic, as most bibliographic databases have different – if any – methods to identify qualitative research[29]. Building upon previous studies[29-31], we developed and piloted a comprehensive search method to identify all articles relevant to our subject, by including not only medical but also psychological bibliographic databases. We combined synonyms of "CKD" with synonyms for "qualitative", "interview", "focus group", "perception", "coping", "barrier", "prognosis", and "preference" to develop search strings for PubMed, Embase, Web of Science, the Cochrane Library, PsycINFO and Emcare. After removal of duplicates, YdJ and EvdW independently selected the relevant titles, abstracts and full-text articles. Review articles and included original articles were checked for missing references (i.e. lateral- or cross-reference searching). Disagreements were resolved by discussion with MvD and YM. Information on the pilot search, the detailed search method, and overview of the search strings and study selection is given in the supplement (**Supplementary material chapter 2, Item 1 and Figure S1**).

2.2.3 Data extraction, quality assessment and synthesis of findings

Data on CKD stage, patient demographics and study methodology were independently extracted by YdJ, EvdW and JM. Correctness of extracting and the accuracy of study characteristics requiring interpretation (e.g. study methodology if not stated by the author) were checked by YdJ and EvdW. Disagreements were resolved by discussion. The COnsolidated criteria for REporting Qualitative research (COREQ), a 32-item checklist[32], was used to assess reporting quality. In line with previous studies[33,34], we categorized studies as having good (≥25 items); moderate (17–24 items); poor (9–16 items); or very poor (≤8 items) reporting quality. A systematic approach following the standards for systematic reviews of qualitative research, as established by the Cochrane Collaboration[33], was used and adapted to our study design. We grouped the included articles in two groups: 1) studies including only patients with CKD, and 2) studies including patients with CKD and other participants, but with sections identifiable as data from patients with CKD. For the first group, all text under 'results' and 'discussion/conclusion' section was used in the analysis; for the second group, only data that could be linked to patients with CKD was extracted. Transcripts were analyzed thematically[34,35]. Articles were read multiple times to familiarize ourselves with the data, and line by line coding of the designated parts was conducted by YdJ, EvdW and JM, summarizing the data using both descriptive and interpretative approaches. Then, by clustering the codes, descriptive themes were identified inductively from the data. As our analysis was inductive, we did not use a predefined or existing coding frame, but developed our own coding frame fitting the data. Coding and analysis was conducted by YdJ, EvdW and JM independently. Hereafter, a discussion on the meaning of the coded text followed in which the coding of the themes was uniformized and the coding tree expanded. After agreement on the coding tree, main themes were created by constant comparison, grouping similar subthemes and organizing subthemes hierarchically into meaningful main themes. To judge consistency of interpretation, themes were compared and discussed. We included and coded all eligible studies. ATLAS.ti software (GmbH, Berlin, version 7.5.18) was used for the coding process.

2.3 Results

2.3.1 Literature search and patient characteristics

Of the 2847 articles identified in the search, 46 studies met the inclusion criteria, representing 1493 participants (**Figure 2.1**). Of these 46 articles, 26 (56%) articles included patients with CKD only, including 529 participants in all CKD stages; the other 20 articles included, amongst others, patients on KRT, conservative therapy, caregivers and healthcare

professionals. An overview of the 26 studies on patients with CKD is given in **Table 2.1**; the remaining 20 studies with mixed populations are presented in **Supplementary material chapter 2, Table S1**. Overall, studies from 12 different countries were included. Although all stages of CKD were included, most studies included participants in CKD stages 3–5.

Figure 2.1 Flowchart of study inclusion. Non-qualitative studies were excluded. Studies that did not contain patients with CKD, or were mixed with other participants and of which the data were not linkable to patients with CKD were marked as 'wrong population'. Studies that did not contain extractable data (e.g. systematic reviews), but were qualitative and included patients with CKD were marked as 'wrong study design'. The inclusion and labelling method is described in more detail in the **Supplementary material, chapter 2**

2.3.2 Synthesis

In total, 4 main themes and 6 subthemes were identified (see **Table 2.2** and **Figure 2.2**).

2.3.2.1 Pursuing normality and control

This theme comprises two subthemes: *pursuing normality* and *a search for knowledge*; both describing the need for certainty.

2.3.2.1.1 Pursuing normality

The gravity of being diagnosed with a life threatening disease, and the realization that with progression of time and decreasing kidney function various outcomes may occur, was unsettling for most: *"Having CKD is just like walking in the valley of the shadow of death, and I can see no hope … My children are still so young. Death has cast a shadow over me, and I am very affected."*[11]. In this disruptive and bewildering period, patients reached a moment where they felt the need to regain control of their disease and return to normality [8,17,36,48,50,52,54,55,60-63]: *"If you can't have some semblance of a normal life, then why would you want to live"*[55]. Especially in the earlier phases of CKD, when few symptoms were experienced, feeling normal instead of feeling like a patient with a chronic disease was relatively easy. However, with the increase of disease severity, participants became more aware of their disease, and expressed an urgency to regain control and stop further deterioration of their health. Patients employed various self-regulation and coping strategies, with searching for information as the main recurring strategy.

Table 2.1 Overview of the 26 studies that included patients with CKD only.

DEMOGRAPHY					CKD STAGE							RESEARCH METHODS			RESULTS
Study	**Country**	**n.**	**Age**	**Sex m/f**	**1**	**2**	**3**	**4**	**5**	**Predialysis**	**Unclear**	**Sampling**	**Data gathering**	**Analysis**	**Principal research aim**
Andrew, J. 2001[43]	United Kingdom	10	-	-						x		Purposive	Interview (semi-structured)	Grounded theory	Needs and experiences of predialysis patients
Iles-Smith, H. 2005[54]	United Kingdom	10	62† (range 37-73)	8/2						x		Consecutive	Interview (semi-structured)	-	Perceptions, expectations and experiences of predialysis patients
Tweed, A. E. 2005[37]	United Kingdom	9	54† (range 29-69)	5/4						x		Convenience	Interview (semi-structured)	Phenomenology	Process of patient decision-making, perspectives and impact on life in predialysis patients
Costantini, L. 2008[9*]	Canada	14	41.3† (range 16-69)	6/8	x	x	x					Purposive	Interview (semi-structured)	Content analysis	Self-management, experiences and perceptions on health and support in CKD
Sakraida, T. J. 2009[59]	USA	6	58† (SD: 9.98)	4/2			x	x				Purposive	Focus group	Ethnography	Perceived resources for and barriers of self-management
Noble, H. 2010[71]	United Kingdom	30	-	-					x			Consecutive	Observations, Interview	-	Symptoms in late stage CKD
de Brito-Ashurst, I. 2011[62]	United Kingdom	20	60† (SD: 8)	0/20							x	Purposive	Focus group, Interview (structured)	-	Views on CKD diets and salt intake and barriers and facilitators of dietary change
Nygardh, A. 2012[52]	Sweden	20	69‡ (range 38-86)	14/6						x		Purposive	Interview (unstructured)	Content analysis	How to empower patients with CKD
Sakraida, T. J. 2012[65]	USA	6	58† (SD: 9.98)	4/2			x					Convenience	Focus group	Ethnography	Coping resources and barriers of self-management in CKD
Walker, R. 2012[48+]	United Kingdom	9	75.9† (range 63-93)	4/5				x	x			Purposive	Interview (semi-structured)	-	Experiences with adherence to behavioural changes in late stage CKD

	DEMOGRAPHY				CKD STAGE							RESEARCH METHODS			RESULTS
Study	**Country**	**n.**	**Age**	**Sex m/f**	**1**	**2**	**3**	**4**	**5**	**Predialysis**	**Unclear**	**Sampling**	**Data gathering**	**Analysis**	**Principal research aim**
Johnston, S. 2012[69]	United Kingdom	9	86† (range 74-96)	4/5					x			Convenience	Interview (semi-structured)	Grounded theory	Motivation to opt for conservative therapy in late-stage CKD
McKillop, G. 2013[68]	United Kingdom	10	60† (range 29-82)	5/5							x	Purposive	Interview (semi-structured)	Thematic analysis	Views and motivations regarding adherence to medication
Lin, C. C. 2013[11]*	Taiwan	15	- (range 25-77)	12/3	x	x	x					Purposive	Interview (semi-structured)	Content analysis	Illness representations and coping processes in early CKD
Lopez-Vargas, P. A. 2014[8]	Australia	38	54† (range 20-79)	23/15	x	x	x	x	x			Purposive	Focus group	Grounded theory	Experiences, perspectives and information needs in managing and living with CKD
Tangkiatkumjai, M. 2014[36]*	Thailand	16	62.5† (SD: 12.3)	6/10			x	x	x			-	Interview (structured)	Thematic analysis	Views on and reasons to use herbal and dietary supplements
Clarke, A. L. 2015[34]	United Kingdom	30	FG: 68.6 (range 48-83); IV 64.1 (range 26-78)	FG: 7/6; IV: 11/6			x	x	x			Convenience, purposive	Focus group, Interview (semi-structured)	Constructivist paradigm	Motivators, barriers and beliefs regarding physical exercise
Erlang, A. S. 2015[57]	Denmark	9	- (range 37-86)	7/2						x		Convenience	Interview (semi-structured)	Phenomenology and hermeneutics	Perspectives, values and experiences related to involvement in the choice of dialysis modality
Shirazian, S. 2016[47]*	USA	23	64† (SD: -)	14/9		x	x	x	x			Purposive	Focus group	Thematic analysis	Views, barriers and supports to the self-management of CKD

	DEMOGRAPHY				CKD STAGE						RESEARCH METHODS			RESULTS
Study	**Country**	**n.**	**Age**	**Sex m/f**	**1**	**2**	**3**	**4**	**5**	**Predialysis** / **Unclear**	**Sampling**	**Data gathering**	**Analysis**	**Principal research aim**
Wright Nunes, J. 2016[64]*	USA	49	62† (SD: 14)	24/25	x	x	x	x	x		Purposive	Interview (semi-structured)	Grounded theory	Emotions after diagnosis, views on how diagnosis was communicated
Wu, C. C. 2016[44]*	Taiwan	15	52† (range 24-81)	7/8				x	x		Purposive	Interview (semi-structured)	Content analysis	Experiences and perceptions related to living with late-stage CKD
Schipper, K. 2016[53]*	The Netherlands	41	- (range 18-75)	17/24			x	x			Purposive	Interview (semi-structured), Focus group	Content analysis	Experience, needs and coping with CKD
Bowling, C. B. 2017[12]+	USA	30	75.1† (range 70.1-90.7)	29/1			x	x	x		Convenience	Focus group	Grounded theory	Self-management and complexity of CKD
Havas, K. 2017[45]*	Australia	63	56.9† (range 25-84)	26/37	x	x	x	x			Convenience	Interview (semi-structured)	Content analysis	Experiences, perceptions and suggestions on self-management support
Lovell, S. 2017[35]	New Zealand	17	75.1† (range 66-90)	14/3					x		Purposive	Interview (semi-structured)	Content analysis	Perspectives of progression of CKD and decision making regarding dialysis.
Pugh-Clarke, K. 2017[58]*	United Kingdom	18	65† (SD: 13.21)	9/9				x	x		Convenience	Interview (semi-structured)	Thematic analysis	Symptom experience in CKD stage 4 and 5
Campbell-Crofts, S. 2018[10]*	Australia	12	- (range 31-81)	4/8			x	x	x		Convenience	Interview (semi-structured)	Thematic analysis	Views on decision making regarding KRT in late stage CKD

Table 2.1 Overview of included studies. Abbreviations: SD: standard deviation, CKD: Chronic Kidney Disease; KRT: kidney replacement therapy. An overview of the 20 studies that included a mixed population is presented in the **Supplementary material chapter 2, Table S1** [a] Studies with unspecified distribution CKD groups (e.g. "CKD 4-5 are marked as CKD 4 and CKD 5); studies with eGFR ranges (e.g. "eGFR <30" are marked as CKD 4 and CKD 5) are attributed to a CKD stage according to the KDIGO CKD staging system; [b] mean; [c] median.

2.3.2.1.2 A search for knowledge

Patients try to gain insight into and understand their disease in order to get a "grip on it". Many patients expressed a great interest in the mechanisms of the disease[8,10,37,41,43,47,51,58,64-67], and methods to avoid further kidney function deterioration. However, this search for information turned out to be a frustrating experience, as patients felt readily accessible and understandable CKD-related information was lacking[38,41,53,66,68,69]. *"I think what is missing from most of these brochures is – WHY? They tell you about it but they don't give you the reason why it's like this."*[8] Appointments with healthcare professionals were often regarded as stressed and hurried[8,9,12,38,40,47,49,51,58,70], and the information received as conflicting[8,12,48,51,58,65], insufficient[9,50,59,65], unclear[38,47,49,61,68,69,71,72], too much[42,58,71], too unspecific and untailored to their situation[9,11,12,40,45,53,62,65,72], or too late[9,44,53]. Subsequently, patients turned to other sources for information, including peers[11,42,45,48,55,61,62,64,67], family members[11,55,59,60,72], friends[55,59,60,61,63,72], online health information[8,9,11,49,66] or mass media[11,63]. Consequently, knowledge on the CKD trajectory towards kidney failure was largely anecdotal, incomplete and not well understood[11,38,63,64,67,72]: *"I have seen my friends go through dialysis and the shows on television. The people on dialysis look so weak and helpless."*[63]. Especially in the absence of symptoms, patients felt no urgency to pursue knowledge on CKD facing these difficulties. However, patients that were content with information provided by healthcare professionals generally felt more empowered, confident, in control, and responsible for their own health[38,40,44,47,49,51,58,65,72]: *"I think it's interesting to know as much as I can about my illness. I mean, the more you know about it, the more chance you've got to influence how it goes – and you're prepared in quite another way for what might happen. That's more or less how I see it."*[47].

2.3.2.2 Prioritizing outcomes

Outcomes prioritized by patients could be grouped into four subthemes (*reaching kidney failure, experienced health, social life*, and *work and economic productivity*), which describe outcomes both directly related to the disease and more personal outcomes.

Theme	Illustrative quotations
Pursuing normality and control	1) Subtheme: pursuing normality a) *"It was with the nurse and she said 'what do you want out of life?' And I said 'I still want to be able to drive and I still want to be able to play golf if possible'. And looking at the [information] booklet she gave me, that [CAPD] looked about the only thing I could do but it's not going to mess my life about any more than I have to. Really trying to keep it at bay. It's there but push it in the corner."*[37] b) *"Yeah, I'm considering peritoneal dialysis because it interferes with your life less. You can do it at night. And it doesn't interfere with your day... If you do it overnight, all your days are free."*[40] c) *"I don't know what it's like to be normal anymore, to feel normal."*[42] 2) Subtheme: a search for knowledge: a) *"The more information I have, the more knowledge I got. That means I can ask better questions, more intelligent questions … otherwise I didn't have a chance to process it."*[59] b) *"(…) [I] shouldn't have to try and read all this medical jargon cause I'm not a—I'm an artist. I'm a painter. I don't know what this means and that means."*[47] c) *"We didn't take 4 years of Latin. An even if we did, it's so far back that we don't remember anymore, and we didn't have medical. So you got to bring down to the level of understanding for the normal person. If it's a kidney, call it a kidney."*[59]
Prioritizing outcomes	1) Subtheme: reaching kidney failure a. *"When they say I've got to go on [dialysis] then I'll work it in, because I've got no choice. It's either that or die." [laughs]*[35] b. *"It was like a monster kind of waiting and lurking in the dark for me and I didn't like the idea at all. Being dependent on the machine for all the functions that you were doing naturally since you were born and the machine takes over and there's no way back. You are not free anymore to make any decisions. If you want to go away it takes so much planning. You are strapped to a machine."*[33] c. *"I'm afraid of receiving dialysis… I want to use everything, which helps me to avoid receiving dialysis."*[36] 2) Subtheme: experienced health a. *"If I'm going to feel this bad for the rest of my life, do I just want to end it now?"* [96] b. *"It's a strange kind of tiredness, quite unlike anything that I've had before. You can't really describe it … it's weird. You just sit down and, phew, you're gone [fallen asleep]. It's weird, strange."*[58] c. *"My thought processes seem to be slowing"*[58] 3) Subtheme: social life a. *"Cultural too, is the male working thing, the identity of working and being a working man, and the stigma of being sick and on dialysis and not being the tough guy"*[51] b. *"The nephrologist is more about making sure the kidney doesn't fail or making sure I live as long as possible, whereas I'm willing to accept some risks for happiness—having a family."*[70] c. *"I can be afraid if I think about the future … Will he still love me if I have more restrictions? And can we stay partners on equal terms?"*[53] 4) Subtheme: work and economic productivity a. *"There's no way I can go back to working where I used to, there's no way I can stand on my feet for 8 hours doing the heavy work I used to do, there's all the retraining and going back into the workforce, plus trying to work out how I'm going to pay my bills, my rent."*[42] b. *"Doing a lot of things that I was able to do six years ago, I can't do that and that's really frustrating, you know, for me because, as my kids know, I worked all my life. I managed a restaurant for 37 years and supported 7 kids … and now I can't work. It's frustrating that I want to go out there and work, do something to help keep me going, and my kids, and I know I can't … Mentally it's like 'Why am I here if I can't do anything to help?'"*[65] c. *"My colleagues and employer don't know that I have CKD. I'm afraid they will use it against me".*[53]

Theme	Illustrative quotations	
Predicting the future	a.	*"He said to me 'Look, you've got a GFR … falling, it is at 22 now which means that you've got about a year left before it's dialysis or transplant.'"*[10]
	b.	*"At the moment he's sitting on the, on his, hands and saying 'Well, it doesn't look like it [dialysis]'ll be happening until sometime next year.'"*[35]
	c.	*"The notion that it will be more difficult in the future is always there. I may not have many problems right now, but the sword of Damocles is always hanging over my head."*[53]
Realising what matters	d.	*"For the last year and half, I've been asking my doctor to change my medications so we can have a child and they keep saying 'next appointment, next appointment.'"* [70]
	e.	*"There is really nothing to discuss with the doctor. […] the doctor is wary and persuaded me to accept dialysis […] all they would do is to encourage me to go on dialysis and tell me the benefits of dialysis."*[41]
	f.	*"He [name of nephrologist] brought up dialysis and was asking me whether I want to have peritoneal dialysis or haemodialysis. During that conversation we seemed to conflict with each other, so what I thought was one thing, he said, 'No, no, no, that's not what you want…' and I'm like 'No, I'm pretty sure I want that'."*[10]

Table 2.2 Overview of the major themes and subthemes with illustrative quotations

2.3.2.2.1 Reaching kidney failure

Although patients prioritized many different outcomes, reaching the moment when KRT initiation would be necessary was predominantly felt as important by most patients[8-11,36,38-44,48,51,53,54,59-62,64,67,68,73,74]: "*(…) I may not have many problems right now, but the sword of Damocles is always hanging over my head*"[53]. It was described as a disastrous, inevitable and constantly looming possibility[8-11,17,37,44,53-55,59-61,64,66,68,74,75]. Choosing for KRT was often expressed as choosing between life and death[11,37,54,73,75], with receiving a kidney transplant seen as the ultimate treatment[39,40,43,48,52,64,76] or, as a patient phrased it: "*(…)the only chance to have a normal life*"[64]. In contrast, dialysis was often regarded as the opposite of quality of life[8,10,38,41,52,54,55,61,63,67,73,75] or an early sign of dying[41,52,59,63,75] while conservative therapy was regarded as choosing for quality of life instead of pointless prolonging[8,54,61,63,75].

2.3.2.2.2 Experienced health

Patients experienced a range of symptoms that were either associated with CKD itself, the disease that caused CKD (e.g. diabetes), treatment side effects, or other comorbidities. In 29 out of 46 articles, a total of 77 different symptoms were mentioned, with the number of symptoms per article ranging between 1 and 50 (presented in **Supplementary material chapter 2, Table S3**). Fatigue and a general feeling of weakness was mentioned by many patients in most articles, although it proved to be difficult to describe to others: *"(…) a feeling, not something obvious. With chronic kidney disease, you don't look different. They tell you, you look good, but they don't see what's inside."*[18]. Fatigue was also often regarded as something normal and consequently, patients felt estrangement and an urge to convince others about the disease severity[48,53,62,65].

2.3.2.2.3 Social life

Living with CKD affected patients' social circles, including family and friends. Some effects were practical in nature, such as burdening family with logistics of CKD and treatment, dialysis preparations or being unable to perform daily tasks[8,10,17,41,43,48,53-55,63,64,73,75,77]. In some cases, these practical concerns influenced decision making, e.g. regarding starting with dialysis[42,63,75]. Living with the consequences of a chronic disease and associated physical and medical restrictions, affected social inclusion and patients' ability to partake in certain social occasions, such as dinner with family or friends[9,12,38,41,43,48,51-54,58,66,70,72]. *"You know, my wife says to me now, you know, we've lost a lot of friends because of my condition, because I've been moody or I get moody, you know. People don't understand what you feel or what you're going through."*[58]. Being unable to fulfil the same role within the family as before CKD was diagnosed caused considerable anxiety amongst patients[11,48,53,59,67]: "*You don't live the life you would like to live. I can't lead the life I envisioned for myself and my kids. (…) I'm just trying to survive.*"[53] Some patients also mentioned the effects of their disease on their sexual wellbeing and family planning[38,47,48,51,53,59,70,73,76]: *"You are not a real man anymore because of your decreased libido. It feels as if I have failed"*[53]. While patients experienced changes in their social role, they also noticed that others changed their behavior towards them. Although patients expressed the desire that their social circles took their symptoms serious, they lamented the social stigma surrounding CKD and felt like they were often solely being regarded as a patient instead of the person they once were[8,48,53,64,67,73]: "*I don't want to have the "stamp" patient, because I don't feel like a patient right now"*[53].

2.3.2.2.4 Work and economic productivity

Being unable to sustain a full-time job resulted in feelings of uselessness. This was emphasized in the absence of clear visible symptoms, and thus legitimization of disease, by their employers or colleagues, as it conflicted with perceived norms of autonomy and sustainability[8,11,39,48,53,59,73]. The effects of CKD on financial independence was often mentioned by patients[8,39,48,53,58,59,63,73] and influenced decision making in some cases[45,48,58,64]: *"With home dialysis, I can work more and support my family and that's really important cause they are reliant on me financially"*[45]. Also, being unable to work was reported to decrease social involvement and acceptance[48,53]: "*Conversations at parties stagnate when you say that you don't work.*"[53].

2.3.2.3 Predicting the future

Patients were interested in both the risk of reaching the outcomes that matter to them, but also the timeframe until these outcomes might occur –indicating that both estimates are important for care planning. Although in some studies patients were given an indication of risk by their healthcare professional[8,10,36,40,51,53,54,68,74], they understood the uncertainty: *"I've*

got a rough timeframe, again its imperfect so no one knows definitively. People say 'When are you going on dialysis?' Well no-one knows but we can guess the way it's going, we can guess"[10]. This left patients in a position where they did not know which outcomes to expect, how high risks for these outcomes were, and when this outcome might occur. This uncertainty regarding their future was accompanied with anxiety, frustration or even resignation to regain control on their disease[8,10,38,40,48,64,65,67,76]. Yet, despite the uncertainty about the future, patients expressed the need to be informed as early as possible on their trajectory nonetheless[9,10,37,44,49,50,53,55,67]: *"The nephrologist advised me not to think about dialysis or transplantation yet because I'm not in that stage of the disease yet. But I know I will need it one day so it's not that easy to put all those emotions and doubts away"*[53].

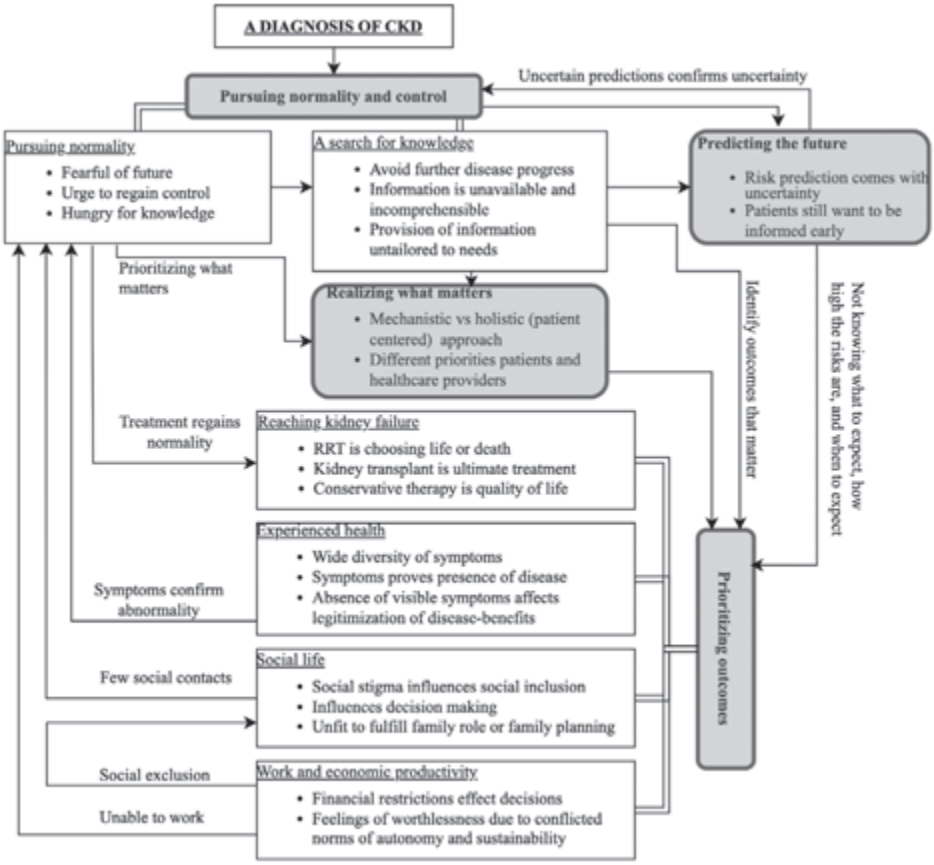


Figure 2.2 Thematic schema: an overview of the identified themes with a hypothesized relation between themes. Patients with CKD face uncertainties and problems regarding their disease progress. This is aggravated by the lack of knowledge, incomprehensible and unavailable information and impossibility to adequately estimate risks, essentially leaving patients in a situation where they do not know what to expect, how high the risks are, and when to expect certain outcomes of interest. Major themes (grey boxes, bold) are linked to subthemes (white boxes, underlined). Abbreviations: CKD; chronic kidney disease, KRT; kidney replacement therapy

2.3.2.4 Realizing what matters

On top of living in a vacuum of prognostic uncertainty, many patients described being misunderstood by their healthcare professionals. Although the exact instances varied widely, there were two main reasons patients felt unheard: 1) professionals displayed a mechanistic approach to disease without an interest in forming relationships, instead of a holistic and person centered approach[9,10,17,40,41,43,47-49,53,65,66,68,75,76]: *"I want to be more than my renal function. They don't see you as a person."*[53], and 2) a difference in priorities between healthcare professionals and patients[8-10,12,37,40,53,54,63,67,68,76,78]: *"My nephrologist just saw kind of being pregnant as an associated risk, not really as a human thing."*[76]. The feeling that not they, but the healthcare professionals were in control of their disease trajectory often resulted in frustration and alienation. Nevertheless, the '*ultimate decision*'[48] whether or not to start KRT was often left in the hands of, or at least influenced heavily by, their healthcare professionals[10,11,40,42,47,51,53,54,61,66,68,76]: *"I am an independent person and I would like to decide about most things. But I also think that if somebody comes and says this is a really bad decision you have made because this, and that and this is supported by arguments then well, I give in to that."*[10].

2.3.3 Comprehensiveness of reporting

The completeness of reporting as assessed by the COREQ was moderate, with studies reporting between 8 to 25 of the 32 items, averaging 18.6 items (58%). A total of four studies scored very good (≥25 items), 30 scored moderate (17–24 items), 11 scored poor (9–16 items) and one scored very poor (≤8 items). Reporting quality was especially weak with regard to describing the domain 'research team and reflexivity' (average 2.8 out of 8), followed by the domain 'study design' (average 9.2 out of 15) and finally, the domain 'analysis and findings' (average 6.5 out of max. 9). A summary of the quality of reporting per domain is shown in **Figure 2.3**. A detailed overview of each study is presented in the **Supplementary material chapter 2, Table S2**.

2.4 DISCUSSION

In this thematic synthesis of 46 qualitative studies, we explored the priorities regarding outcomes of patients with CKD and barriers encountered regarding person-centered care and care planning. The themes that emerged describe the health journey after diagnosis with CKD, underline the disruptiveness of CKD on all aspects of life, and the urgency felt for incorporation of person-centered care within routine medical care. We identified four major themes with six subthemes: *pursuing normality and control* (subthemes: *pursuing normality*; *a search for knowledge*); *prioritizing outcomes* (subthemes: *reaching kidney failure; experienced health; social life; work and economic productivity*); *predicting the future* and *realizing what matters*. Three barriers relevant to person-centered care provision were embedded within these

themes: untailored and incomprehensible information, the inability to accurately estimate risks, and differences in priorities regarding outcomes and care processes between patients and healthcare professionals. The overall completeness of reporting as assessed by the COREQ was moderate, especially so for the domain 'research team and reflexivity'.

Figure 2.3 COREQ quality of reporting summary of the 46 included studies, over the three domains (upper panel: domain 1 'research team and reflexivity', comprises 8 signaling questions which describes both the personal characteristics of the researchers and their relationships with the participants; middle panel: domain 2 'study design' comprises 15 questions which describes the included population and study methods; and lower panel: domain 3 'analysis and findings', comprises 9 questions which describes the analysis and clarity of the results) containing a total of 32 signaling questions. An overview of each individual study is presented in the **Supplementary material chapter 2, Table S2**

2.4.1 Key findings

Following the disrupting period after diagnosis of CKD, patients express the need to return to normality and regain control to avoid further deterioration of kidney function and associated physical and mental symptoms. However, as patients in early stages of CKD usually experience few symptoms, the initial shock of being diagnosed with a chronic disease subdues, and maintaining or ignoring the status quo turned out to be relatively easy. A complicating factor in regaining control was clearly described by many patients in our study, namely the struggle and frustration to gain comprehensible information tailored to their specific situation, which we identified as an important barrier for person centered care provision. As a consequence of both the absence of symptoms and the difficulties in obtaining relevant information, self-management strategies were postponed by patients. The delayed self-management activation but also the strive for normality in the earlier stages of disease are not unique to CKD, but are observed in other chronic diseases as well[79,80].

As time passed and disease progresses, patients with CKD learn about, or in some instances already experience some of the possible outcomes related to CKD - both directly related to the disease, such as physical or mental symptoms, but also indirectly, such as a social stigma or financial burdens. Prioritization of these outcomes differed greatly between the patients in our study, but one outcome was emphasized and feared most: reaching kidney failure and choosing between dialysis, transplant or conservative care. Although discussing this topic is regarded as a difficult subject, both by clinicians and patients, it is often recognized as an important subject and thus prioritized and facilitated by healthcare professionals. This is in contrast to the other three groups of outcomes which were regarded by some patients as equally important: apart from kidney failure, patients worry about the symptoms associated with CKD – both physical and mental symptoms, the effects of CKD on their social life and on economic productivity. These other aspects of disease are not routinely assessed by healthcare professionals and often go unnoticed as a result. Consequently, patients feel misunderstood by their healthcare professionals, as they realize that their priorities do not match those of their healthcare professionals. Indeed, many patients in our study expressed the need for holistic care, instead of an approach they described as mechanistic: a focus on laboratory results instead of their actual perceived wellbeing. This barrier for person centered care was mentioned in most studies, and caused considerable frustration with, and alienation from healthcare professionals. Although healthcare professionals are aware of the disruptive effects of CKD on these important aspects of life[18,81], traditionally the main focus of care is on prolongation of time to kidney failure or death[81]. Illustrative, in a 2003 survey, US nephrology fellows

reported that palliative care training was not integrated sufficiently in their curricula, and consequently they felt unprepared to discuss end-of-life issues[82]. Despite that, and even though the majority of our included articles have been published in between, a repeat survey 10 years later showed similar results[83].

Patients have a realistic expectation that neither the risks of future outcomes nor the timeframe of reaching them can be predicted with a large degree of certainty. This realization causes feelings of anxiety and frustration, and consequently, we identified this as the third and final barrier for person centered care implementation. Yet, despite the uncertainty of the risk estimates, participants in our study expressed eagerness to be informed as early as possible, and urged not to withhold information on prognosis. We identified many instances where selective, delayed or incomplete information on sensitive topics such as disease progression or kidney failure frequently resulted in frustration and in some cases even mistrust. Clinicians are aware of this prognostic uncertainty, but refrain from discussing risks because they lack aids to adequately counsel patients on the outcomes of their interest[84], or fear to emotionally overwhelm patients[81,85]. However, deciding early and planning in advance which treatment option is most suitable or which outcomes to avoid, has been shown to positively enhance patients' coping with disease[86], especially when the preferences of patients are taken into account[87]. This relation between risk uncertainty and focus on kidney failure or prolonging survival is illustrated for example by the number of prognostic prediction models that have been developed for these outcomes in patients with CKD: for KRT and death respectively 42 and 16 models were identified in systematic reviews[88,89], and models validated in these populations perform poorly[90-92]. In contrast, no models for other outcomes prioritized by patients exist. Interestingly, contrasting the number of prediction models on this topic, the risk of death was mentioned only a few times by patients, usually in the context as an alternative for KRT, i.e. conservative treatment[18].

2.4.2 Implications

Our study provides several clinical implications. Patients were frustrated about the lack of available and accessible information, and realize that disease education is not prioritized by their healthcare professionals. Consequently, they look for information elsewhere, resulting in incomplete or even incorrect information. Several systematic reviews on patient education and self-management have demonstrated positive effects of education on knowledge and self-management, though the number of included studies was low and the effects dependent on the type of educational interventions and setting[79,93,94]. Studies in other chronic diseases, such as diabetes[95,96] and hypertension[94,97] demonstrated similar results. Our findings thus underline the importance of disease education in CKD. Next, patients with CKD describe

the wide array of problems they experience, but feel unheard and misunderstood by their healthcare professionals. For example, one of the recurring themes was the influence of CKD on work and economic productivity. Conversations between healthcare professionals and patients might stimulate that healthcare professionals support patients in coping with work related concerns, make appropriate referrals to a social worker, or help patients arranging a more flexible work environment. Another finding was the struggle with a wide array of disease related physical and mental symptoms of CKD - we identified a total of 77 distinct symptoms -, which often remained unnoticed by their healthcare professionals. Our findings could thus serve as a guide for identification of care needs for healthcare professionals. More formally, our study could be used as a starting point in the development or selection of PROMs and incorporation of these PROMs within routine CKD care[57]. Incorporation of person-centered care and PROMs in CKD shows promising results[23-25] and may result in outcomes that are more satisfactory[98]. Finally, patients realize that the prediction of outcomes of interest is largely impossible. Prediction studies on the development of kidney failure or the risk of death have been conducted, however these cover only a small part of the spectrum of outcomes that are important to patients. Future prediction studies could focus on other patient-prioritized outcomes (for examples predicting outcomes such as 'when will I have to give up work?' or 'when will I be unable to drive?') or on predicting PROs: similar studies have been conducted in orthopedics[99-102], neurosurgery[103], and psychiatry[104].

2.4.3 Strengths and limitations

Our study comes with strengths and weaknesses. This is the first study to comprehensively provide an overview of outcomes prioritized by patients with CKD and barriers for person centered care provision by means of a systematic review of qualitative studies, using a broad scope by not focusing on the medical side alone. Thematic synthesis of qualitative studies instead of original data is a relatively novel method to achieve abstraction and transferability at a higher level beyond the included original studies[22]. Another strength of this study is the inclusion of a large number of patients in all stages of CKD, from many different countries including a diverse demographic and many different ethnicities. Our study is however not without its limitations. First, without inclusion of non-English articles, transferability to non-English speaking populations is unclear, although we included several articles with quotes that were translated to English. As indicated with the COREQ criteria, most of the included studies incompletely reported information on their methodology or findings, which may have impacted the validity of our results. As this is not uncommon in qualitative research[105], and because the aim of our study was to ensure a broad range of perspectives were captured and to encourage transparency and

transferability of findings, we included all studies regardless. Next, because most studies were conducted in developed countries, the transferability of our findings to developing countries is uncertain. Finally, as with all qualitative research, interpretation and reporting of findings is influenced by the personal beliefs and biases of the researcher (i.e. research reflexivity). To prevent that data interpretation and results were strongly colored by the preconceptions of a single profession, we purposely created a team of authors with a wide diversity of professional background and experience.

2.5 Conclusion

Living with a diagnosis of CKD has a major impact not only on physical outcomes, but on many other aspects as well, such as mental health, social life and emotional wellbeing. Inadequate provision of information tailored to both the stage of the disease and the capacities of the patient, uncertainty regarding the prognosis and difference in priorities between healthcare professionals and patients are barriers that stand in the way to optimal person-centered healthcare. Multidisciplinary care and regular use of PROMs in nephrological care may be a strategy to help focus care on the needs and outcomes of most importance to adults with CKD.

2.5 REFERENCES

1. Slevin J TA. Understanding what the public know about their kidneys and what they do. Findings from Ipsos MORI survey 2014 [Available from: https://www.thinkkidneys.nhs.uk/wp-content/uploads/2015/01/ThinkKidneys-Report-270115-Understanding-what-the-public-know-about-theirkidneys-and-what-they-do.pdf accessed 22-02-2018.
2. Coresh J, Byrd-Holt D, Astor BC, et al. Chronic kidney disease awareness, prevalence, and trends among U.S. adults, 1999 to 2000. J Am Soc Nephrol 2005;16(1):180–8. doi: https://doi.org/10.1681/asn.2004070539 [published Online First: 2004/11/26].
3. Wright JA, Wallston KA, Elasy TA, et al. Development and results of a kidney disease knowledge survey given to patients with CKD. Am J Kidney Dis 2011;57(3):387–95. doi: https://doi.org/10.1053/j.ajkd.2010.09.018 [published Online First: 2010/12/21].
4. Finkelstein FO, Story K, Firanek C, et al. Perceived knowledge among patients cared for by nephrologists about chronic kidney disease and end-stage renal disease therapies. Kidney Int 2008;74(9):1178–84. doi: https://doi.org/10.1038/ki.2008.376 [published Online First: 2008/08/01].
5. Mehrotra R, Marsh D, Vonesh E, et al. Patient education and access of ESRD patients to renal replacement therapies beyond in-center hemodialysis. Kidney Int 2005;68(1):378–90. doi: https://doi.org/10.1111/j.1523-1755.2005.00453.x [published Online First: 2005/06/16].
6. Tuot DS, Zhu Y, Velasquez A, et al. Variation in Patients' Awareness of CKD according to How They Are Asked. Clin J Am Soc Nephrol 2016;11(9):1566–73. doi: https://doi.org/10.2215/cjn.00490116 [published Online First: 2016/06/25].
7. Tuot DS, Plantinga LC. What patients don't know may hurt them: knowledge and the perception of knowledge among patients with CKD. Kidney Int 2011;80(12):1256–7. doi: https://doi.org/10.1038/ki.2011.269 [published Online First: 2011/12/01].
8. Lopez-Vargas PA, Tong A, Phoon RK, et al. Knowledge deficit of patients with stage 1–4 CKD: a focus group study. Nephrology (Carlton) 2014;19(4):234–43. doi: https://doi.org/10.1111/nep.12206 [published Online First: 2014/01/17].
9. Costantini L, Beanlands H, McCay E, et al. The self-management experience of people with mild to moderate chronic kidney disease. Nephrol Nurs J 2008;35(2):147–55; quiz 56. [published Online First: 2008/05/14].
10. Campbell-Crofts S, Stewart G. How perceived feelings of "wellness" influence the decision-making of people with predialysis chronic kidney disease. J Clin Nurs 2018;27(7–8):1561–71. doi: https://doi.org/10.1111/jocn.14220 [published Online First: 2017/12/15].
11. Lin CC, Chen MC, Hsieh HF, et al. Illness representations and coping processes of Taiwanese patients with early-stage chronic kidney disease. J Nurs Res 2013;21(2):120–8. doi: https://doi.org/10.1097/jnr.0b013e3182921fb8 [published Online First: 2013/05/18].
12. Bowling CB, Vandenberg AE, Phillips LS, et al. Older Patients' Perspectives on Managing Complexity in CKD Self-Management. Clin J Am Soc Nephrol 2017;12(4):635–43. doi: https://doi.org/10.2215/cjn.06850616 [published Online First: 2017/04/09].
13. Aiyegbusi OL, Kyte D, Cockwell P, et al. A patient-centred approach to measuring quality in kidney care: patient-reported outcome measures and patient-reported experience measures. Current opinion in nephrology and hypertension 2017;26(6):442–49. doi: https://doi.org/10.1097/mnh.0000000000000357 [published Online First: 2017/08/15].
14. de Goeij MCM, Ocak G, Rotmans JI, Eijgenraam JW, Dekker FW, Halbesma N. Course of symptoms and health-related quality of life during specialized pre-dialysis care. PLoS One. 2014;9(4):e93069–e69. https://doi.org/10.1371/journal.pone.0093069.

15. Silés-González J, Solano-Ruiz M. Qualitative research and narrative sources in the context of critical and renal cares. Open J Nurs. 2014;4(06):385–401. https://doi.org/10.4236/ojn.2014.46045.

16. Bruce A, Sheilds L, Molzahn A, et al. Stories of liminality: living with life-threatening illness. J Holistic Nurs 2014;32(1):35–43. doi: https://doi.org/10.1177/0898010113498823 [published Online First: 2013/08/09].

17. Tong A, Sainsbury P, Carter SM, et al. Patients' priorities for health research: focus group study of patients with chronic kidney disease. Nephrol DialysisTranspl 2008;23(10):3206–14. doi: https://doi.org/10.1093/ndt/gfn207 [published Online First: 2008/05/01].

18. Ramer SJ, McCall NN, Robinson-Cohen C, et al. Health Outcome Priorities of Older Adults with Advanced CKD and Concordance with Their Nephrology Providers' Perceptions. J Am Soc Nephrol 2018;29(12):2870–78. doi: https://doi.org/10.1681/asn.2018060657 [published Online First: 2018/11/06].

19. Breckenridge K, Bekker HL, Gibbons E, et al. How to routinely collect data on patient-reported outcome and experience measures in renal registries in Europe: an expert consensus meeting. Nephrol DialysisTranspl 2015;30(10):1605–14. doi: https://doi.org/10.1093/ndt/gfv209 [published Online First: 2015/05/20].

20. Anderson NE, Calvert M, Cockwell P, et al. The Use of Patient-Reported Outcomes in Patients Treated With Maintenance Hemodialysis: A Perspective. Am J Kidney Dis 2019;74(3):399–406. doi: https://doi.org/10.1053/j.ajkd.2019.01.035 [published Online First: 2019/04/30].

21. van der Veer SN, Aresi G, Gair R. Incorporating patient-reported symptom assessments into routine care for people with chronic kidney disease. Clin Kidney J 2017;10(6):783–87. doi: https://doi.org/10.1093/ckj/sfx106 [published Online First: 2017/12/19].

22. Dixon-Woods M, Agarwal S, Jones D, et al. Synthesising qualitative and quantitative evidence: a review of possible methods. J Health Serv Res Policy 2005;10(1):45–53. doi: https://doi.org/10.1177/135581960501000110 [published Online First: 2005/01/26].

23. Bingham CO, 3rd, Noonan VK, Auger C, et al. Montreal Accord on Patient-Reported Outcomes (PROs) use series - Paper 4: patient-reported outcomes can inform clinical decision making in chronic care. J Clin Epidemiol 2017;89:136–41. doi: https://doi.org/10.1016/j.jclinepi.2017.04.014 [published Online First: 2017/04/24].

24. Calvert M, Kyte D, Price G, et al. Maximising the impact of patient reported outcome assessment for patients and society. BMJ (Clinical research ed) 2019;364:k5267. doi: https://doi.org/10.1136/bmj.k5267 [published Online First: 2019/01/27].

25. Morton RL, Sellars M. From patient-centered to person-centered Care for Kidney Diseases. Clin J Am Soc Nephrol. 2019;14(4):623–5. https://doi.org/10.2215/cjn.10380818.

26. Tong A, Flemming K, McInnes E, et al. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. BMC Med Res Methodol 2012;12:181. doi: https://doi.org/10.1186/1471-2288-12-181 [published Online First: 2012/11/29].

27. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ (Clinical research ed) 2009;339:b2535. doi: https://doi.org/10.1136/bmj.b2535

28. Smith HJ, Chen J, Liu X. Language and rigour in qualitative research: problems and principles in analyzing data collected in mandarin. BMC Med Res Methodol. 2008;8(1):44. https://doi.org/10.1186/1471-2288-8-44.

29. Walters LA, Wilczynski NL, Haynes RB. Developing optimal search strategies for retrieving clinically relevant qualitative studies in EMBASE. Qual Health Res 2006;16(1):162–8. doi: https://doi.org/10.1177/1049732305284027 [published Online First: 2005/12/01].

30. Barroso J, Gollop CJ, Sandelowski M, et al. The challenges of searching for and retrieving qualitative studies. West J Nurs Res 2003;25(2):153–78. doi: https://doi.org/10.1177/0193945902250034 [published Online First: 2003/04/02].

31. Finfgeld-Connett D, Johnson ED. Literature search strategies for conducting knowledge-building and theory-generating qualitative systematic reviews. J Adv Nurs 2013;69(1):194–204. doi: https://doi.org/10.1111/j.1365-2648.2012.06037.x [published Online First: 2012/05/18].

32. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. International journal for quality in health care : J Int Soc Quality Health Care 2007;19(6):349–57. doi: https://doi.org/10.1093/intqhc/mzm042 [published Online First: 2007/09/18].

33. Hannes K. Critical appraisal of qualitative research. In: Noyes J BA, Hannes K, Harden A, Harris J, Lewin S, Lockwood C ed. Supplementary Guidance for Inclusion of Qualitative Research in Cochrane Systematic Reviews of Interventions: Cochrane Collaboration Qualitative Methods Group 2011.

34. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol. 2006;3(2):77–101. https://doi.org/10.1191/1478088706qp063oa.

35. Braun V, Clarke V. Thematic analysis. APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological. Washington, DC, US: American Psychological Association 2012:57–71.

36. Boulware LE, Hill-Briggs F, Kraus ES, et al. Identifying and addressing barriers to African American and non-African American families' discussions about preemptive living related kidney transplantation. Prog Transpl 2011;21(2):97–104; quiz 05. [published Online First: 2011/07/09].

37. Lederer S, Fischer MJ, Gordon HS, et al. A question prompt sheet for adult patients with chronic kidney disease. BMC Nephrol 2016;17(1):155. doi: https://doi.org/10.1186/s12882-016-0362-z [published Online First: 2016/10/21].

38. Iles-Smith H. Perceptions and experiences of pre-dialysis patients. EDTNA/ERCA journal (English ed) 2005;31(3):130–3. [published Online First: 2005/12/21].

39. Beard BH. Fear of death and fear of life. The dilemma in chronic renal failure, hemodialysis, and kidney transplantation. Arch Gen Psychiatry 1969;21(3):373–80. [published Online First: 1969/09/01].

40. Erlang AS, Nielsen IH, Hansen HO, et al. Patients experiences of involvement in choice of dialysis mode. J Ren Care 2015;41(4):260–7. doi: https://doi.org/10.1111/jorc.12141 [published Online First: 2015/09/30].

41. Shirazian S, Crnosija N, Weinger K, et al. The self-management experience of patients with type 2 diabetes and chronic kidney disease: A qualitative study. Chronic illness 2016;12(1):18–28. doi: https://doi.org/10.1177/1742395315614381 [published Online First: 2015/11/06].

42. Loiselle m-c, Michaud C, M. O'Connor A. Decisional Needs Assessment to Help Patients with Advanced Chronic Kidney Disease Make Better Dialysis Choices2016.

43. Andrew J. The pre-dialysis experience--are individual needs being met? EDTNA/ERCA journal (English ed) 2001;27(2):72–4. [published Online First: 2002/03/01].

44. Wright Nunes J, Roney M, Kerr E, et al. A diagnosis of chronic kidney disease: despite fears patients want to know early. Clin Nephrol 2016;86(2):78–86. doi: https://doi.org/10.5414/cn108831 [published Online First: 2016/06/28].

45. Walker RC, Howard K, Morton RL, et al. Patient and caregiver values, beliefs and experiences when considering home dialysis as a treatment option: a semi-structured interview study. Nephrol DialysisTranspl 2016;31(1):133–41. doi: https://doi.org/10.1093/ndt/gfv330 [published Online First: 2015/09/09].

46. Al-Moghrabi D, Tsichlaki A, Alkadi S, et al. How well are dental qualitative studies involving interviews and focus groups reported? Journal of Dentistry 2019;84:44–48. doi: https://doi.org/10.1016/j.jdent.2019.03.001.

47. Nygardh A, Malm D, Wikby K, et al. The experience of empowerment in the patient-staff encounter: the patient's perspective. J Clin Nurs 2012;21(5–6):897–904. doi: https://doi.org/10.1111/j.1365-2702.2011.03901.x [published Online First: 2011/11/16].

48. Tong A, Sainsbury P, Chadban S, et al. Patients' experiences and perspectives of living with CKD. Am J Kidney Dis 2009;53(4):689–700. doi: https://doi.org/10.1053/j.ajkd.2008.10.050 [published Online First: 2009/02/14].

49. Sakraida TJ, Robinson MV. Health literacy self-management by patients with type 2 diabetes and stage 3 chronic kidney disease. West J Nurs Res 2009;31(5):627–47. doi: https://doi.org/10.1177/0193945909334096 [published Online First: 2009/04/25].

50. Lee A, Gudex C, Povlsen JV, et al. Patients' views regarding choice of dialysis modality. Nephrol DialysisTranspl 2008;23(12):3953–9. doi: https://doi.org/10.1093/ndt/gfn365 [published Online First: 2008/07/01].

51. Walker R, James H, Burns A. Adhering to behaviour change in older pre-dialysis populations--what do patients think? A qualitative study. J Ren Care 2012;38(1):34–42. doi: https://doi.org/10.1111/j.1755-6686.2012.00262.x [published Online First: 2012/03/01].

52. Morton RL, Tong A, Webster AC, et al. Characteristics of dialysis important to patients and family caregivers: a mixed methods approach. Nephrol DialysisTranspl 2011;26(12):4038–46. doi: https://doi.org/10.1093/ndt/gfr177 [published Online First: 2011/04/13].

53. Schipper K, van der Borg WE, de Jong-Camerik J, et al. Living with moderate to severe renal failure from the perspective of patients. BMC Nephrol 2016;17(1):48. doi: https://doi.org/10.1186/s12882-016-0263-1 [published Online First: 2016/05/18].

54. Lovell S, Walker RJ, Schollum JB, et al. To dialyse or delay: a qualitative study of older New Zealanders' perceptions and experiences of decision-making, with stage 5 chronic kidney disease. BMJ Open 2017;7(3):e014781. doi: https://doi.org/10.1136/bmjopen-2016-014781 [published Online First: 2017/04/01].

55. Tweed AE, Ceaser K. Renal replacement therapy choices for pre-dialysis renal patients. Br J Nurs 2005;14(12):659–64. doi: https://doi.org/10.12968/bjon.2005.14.12.18287 [published Online First: 2005/07/13].

56. Walsh S, Jones M, Bressington D, McKenna L, Brown E, Terhaag S, et al. Adherence to COREQ reporting guidelines for qualitative research: a Scientometric study in nursing social science. Int J Qual Methods. 2020;19:1609406920982145. https://doi.org/10.1177/1609406920982145.

57. van der Willik EM, Hemmelder MH, Bart HAJ, van Ittersum FJ, Hoogendijk-van den Akker JM, Bos WJW, et al. Routinely measuring symptom burden and health-related quality of life in dialysis patients: first results from the Dutch registry of patient-reported outcome measures. Clin Kidney J. 2020;14(6):1535–44. https://doi.org/10.1093/ckj/sfz192.

58. Lo C, Ilic D, Teede H, et al. The Perspectives of Patients on Health-Care for Co-Morbid Diabetes and Chronic Kidney Disease: A Qualitative Study. PLoS One 2016;11(1):e0146615. doi: https://doi.org/10.1371/journal.pone.0146615 [published Online First: 2016/01/06].

59. Sakraida TJ, Robinson MV. Mental health and relational self-management experiences of patients with type 2 diabetes and stage 3 chronic kidney disease. Issues Ment Health Nurs 2012;33(11):786–96. doi: https://doi.org/10.3109/01612840.2012.713446 [published Online First: 2012/11/14].

60. Tangkiatkumjai M, Boardman H, Praditpornsilpa K, et al. Reasons why Thai patients with chronic kidney disease use or do not use herbal and dietary supplements. BMC Complement Altern Med 2014;14:473. doi: https://doi.org/10.1186/1472-6882-14-473 [published Online First: 2014/12/08].

61. Llewellyn H, Low J, Smith G, et al. Narratives of continuity among older people with late stage chronic kidney disease who decline dialysis. Soc Sci Med (1982) 2014;114:49–56. doi: https://doi.org/10.1016/j.socscimed.2014.05.037 [published Online First: 2014/06/10].

62. Clarke AL, Young HM, Hull KL, et al. Motivations and barriers to exercise in chronic kidney disease: a qualitative study. Nephrol Dialysis Transpl 2015;30(11):1885–92. doi: https://doi.org/10.1093/ndt/gfv208 [published Online First: 2015/06/10].

63. Seah AS, Tan F, Srinivas S, et al. Opting out of dialysis - Exploring patients' decisions to forego dialysis in favour of conservative non-dialytic management for end-stage renal disease. Health Expectations 2015;18(5):1018–29. doi: https://doi.org/10.1111/hex.12075 [published Online First: 2013/05/08].

64. Wu CC, Lin CC, Hsieh HF, et al. Lived experiences and illness representation of Taiwanese patients with late-stage chronic kidney disease. J Health Psychol 2016;21(12):2788–98. doi: https://doi.org/10.1177/1359105315587134 [published Online First: 2015/06/11].

65. Havas K, Douglas C, Bonner A. Closing the loop in person-centered care: patient experiences of a chronic kidney disease self-management intervention. Patient Prefer Adherence 2017;11:1963–73. doi: https://doi.org/10.2147/ppa.s147831 [published Online First: 2017/12/15].

66. Meuleman Y, Ten Brinke L, Kwakernaak AJ, et al. Perceived Barriers and Support Strategies for Reducing Sodium Intake in Patients with Chronic Kidney Disease: a Qualitative Study. Int J Behav Med 2015;22(4):530–9. doi: https://doi.org/10.1007/s12529-014-9447-x [published Online First: 2014/10/10].

67. Walker RC, Walker S, Morton RL, et al. Maori patients' experiences and perspectives of chronic kidney disease: a New Zealand qualitative interview study. BMJ Open 2017;7(1):e013829. doi: https://doi.org/10.1136/bmjopen-2016-013829 [published Online First: 2017/01/21].

68. Kang H, Stenfors-Hayes T. Feeling Well and Having Good Numbers: Renal Patients' Encounter With Clinical Uncertainties and the Responsibility to "Live Well". Qual Health Res 2016;26(12):1591–602. doi: https://doi.org/10.1177/1049732315591484 [published Online First: 2015/07/02].

69. Lederer S, Fischer MJ, Gordon HS, et al. Barriers to effective communication between veterans with chronic kidney disease and their healthcare providers. Clin Kidney J 2015;8(6):766–71. doi: https://doi.org/10.1093/ckj/sfv079 [published Online First: 2015/11/28].

70. Pugh-Clarke K, Read SC, Sim J. Symptom experience in non-dialysis-dependent chronic kidney disease: A qualitative descriptive study. J Ren Care 2017;43(4):197–208. doi: https://doi.org/10.1111/jorc.12208 [published Online First: 2017/06/15].

71. Smekal MD, Tam-Tham H, Finlay J, et al. Perceived Benefits and Challenges of a Risk-Based Approach to Multidisciplinary Chronic Kidney Disease Care: A Qualitative Descriptive Study. Can J Kidney Health Dis 2018;5:2054358118763809. doi: https://doi.org/10.1177/2054358118763809 [published Online First: 2018/04/11].

72. de Brito-Ashurst I, Perry L, Sanders TA, et al. Barriers and facilitators of dietary sodium restriction amongst Bangladeshi chronic kidney disease patients. J Hum Nutr Diet 2011;24(1):86–95. doi: https://doi.org/10.1111/j.1365-277X.2010.01129.x [published Online First: 2010/12/01].

73. Subramanian L, Quinn M, Zhao J, et al. Coping with kidney disease - qualitative findings from the Empowering Patients on Choices for Renal Replacement Therapy (EPOCH-RRT) study. BMC Nephrol 2017;18(1):119. doi: https://doi.org/10.1186/s12882-017-0542-5 [published Online First: 2017/04/05].

74. McKillop G, Joy J. Patients' experience and perceptions of polypharmacy in chronic kidney disease and its impact on adherent behaviour. J Ren Care 2013;39(4):200–7. doi: https://doi.org/10.1111/j.1755-6686.2013.12037.x [published Online First: 2013/11/20].

75. Johnston S, Noble H. Factors influencing patients with stage 5 chronic kidney disease to opt for conservative management: a practitioner research study. J Clin Nurs 2012;21(9–10):1215–22. doi: https://doi.org/10.1111/j.1365-2702.2011.04001.x [published Online First: 2012/03/06].

76. Tong A, Brown MA, Winkelmayer WC, et al. Perspectives on Pregnancy in Women With CKD: A Semistructured Interview Study. Am J Kidney Dis 2015;66(6):951–61. doi: https://doi.org/10.1053/j.ajkd.2015.08.023 [published Online First: 2015/10/11].

77. Noble H, Meyer PJ, Bridge DJ, et al. Exploring symptoms in patients managed without dialysis: a qualitative research study. J Ren Care 2010;36(1):9–15. doi: https://doi.org/10.1111/j.1755-6686.2010.00140.x [published Online First: 2010/03/11].

78. Schipper K, Landeweer E, Abma TA. Living with end-stage renal disease. Nurs Ethics 2017:969733016687154. doi: https://doi.org/10.1177/0969733016687154 [published Online First: 2017/01/18], 2018.

79. Lopez-Vargas PA, Tong A, Howell M, et al. Educational Interventions for Patients With CKD: A Systematic Review. Am J Kidney Dis 2016;68(3):353–70. doi: https://doi.org/10.1053/j.ajkd.2016.01.022 [published Online First: 2016/03/30].

80. Lopez-Vargas PA, Tong A, Howell M, et al. Patient awareness and beliefs about the risk factors and comorbidities associated with chronic kidney disease : A mixed-methods study. Nephrology (Carlton) 2017;22(5):374–81. doi: https://doi.org/10.1111/nep.12829 [published Online First: 2016/06/02].

81. Grubbs V, Tuot DS, Powe NR, et al. System-Level Barriers and Facilitators for Foregoing or Withdrawing Dialysis: A Qualitative Study of Nephrologists in the United States and England. Am J Kidney Dis 2017;70(5):602–10. doi: https://doi.org/10.1053/j.ajkd.2016.12.015 [published Online First: 2017/03/01].

82. Holley JL, Carmody SS, Moss AH, et al. The need for end-of-life care training in nephrology: national survey results of nephrology fellows. Am J Kidney Dis 2003;42(4):813–20. doi: https://doi.org/10.1016/s0272-6386(03)00868-0 [published Online First: 2003/10/02].

83. Combs SA, Culp S, Matlock DD, et al. Update on end-of-life care training during nephrology fellowship: a cross-sectional national survey of fellows. Am J Kidney Dis 2015;65(2):233–9. doi: https://doi.org/10.1053/j.ajkd.2014.07.018 [published Online First: 2014/09/24].

84. Dowen F, Sidhu K, Broadbent E, Pilmore H Communicating projected survival with treatments for chronic kidney disease: patient comprehension and perspectives on visual aids. BMC Med Inform Decis Mak 2017; 17(1). http://europepmc.org/abstract/MED/28934951. https://doi.org/10.1186/s12911-017-0536-z. https://europepmc.org/articles/PMC5607842. https://europepmc.org/articles/PMC5607842?pdf=render (accessed 2017/09//).

85. Greer RC, Crews DC, Boulware LE. Challenges perceived by primary care providers to educating patients about chronic kidney disease. J Ren Care 2012;38(4):174–81. doi: https://doi.org/10.1111/j.1755-6686.2012.00323.x [published Online First: 2012/11/28].

86. Davison SN, Simpson C. Hope and advance care planning in patients with end stage renal disease: qualitative interview study. BMJ (Clinical research ed) 2006;333(7574):886. doi: https://doi.org/10.1136/bmj.38965.626250.55 [published Online First: 2006/09/23].

87. Eraker SA, Kirscht JP, Becker MH. Understanding and improving patient compliance. Ann Intern Med 1984;100(2):258–68. doi: https://doi.org/10.7326/0003-4819-100-2-258 [published Online First: 1984/02/01].

88. Ramspek CL, de Jong Y, Dekker FW, et al. Towards the best kidney failure prediction tool: a systematic review and selection aid. Nephrol DialysisTranspl 2019 doi: https://doi.org/10.1093/ndt/gfz018 [published Online First: 2019/03/05].

89. Ramspek CL, Voskamp PW, van Ittersum FJ, et al. Prediction models for the mortality risk in chronic dialysis patients: a systematic review and independent external validation study. Clin Epidemiol 2017;9:451–64. doi: https://doi.org/10.2147/clep.S139748 [published Online First: 2017/09/19].

90. de Jong Y, Fu EL, van Diepen M, Trevisan M, Szummer K, Dekker FW, et al. Validation of risk scores for ischaemic stroke in atrial fibrillation across the spectrum of kidney function. Eur Heart J. 2021;42(15):1476–85. https://doi.org/10.1093/eurheartj/ehab059.

91. de Jong Y, Ramspek CL, van der Endt VHW, et al. A systematic review and external validation of stroke prediction models demonstrates poor performance in dialysis patients. J Clin Epidemiol 2020;123:69–79. doi: https://doi.org/10.1016/j.jclinepi.2020.03.015.

92. de Jong Y, Ramspek CL, Zoccali C, Jager KJ, Dekker FW, van Diepen M. Appraising prediction research: a guide and meta-review on bias and applicability assessment using the Prediction model Risk Of Bias ASsessment Tool (PROBAST). Nephrology. https://doi.org/10.1111/nep.13913.

93. Galbraith L, Jacobs C, Hemmelgarn BR, et al. Chronic disease management interventions for people with chronic kidney disease in primary care: a systematic review and meta-analysis. Nephrol DialysisTranspl 2018;33(1):112–21. doi: https://doi.org/10.1093/ndt/gfw359 [published Online First: 2017/01/18].

94. Zimbudzi E, Lo C, Misso ML, et al. Effectiveness of self-management support interventions for people with comorbid diabetes and chronic kidney disease: a systematic review and meta-analysis. Syst Rev 2018;7(1):84. doi: https://doi.org/10.1186/s13643-018-0748-z [published Online First: 2018/06/15].

95. Pimouguet C, Le Goff M, Thiébaut R, et al. Effectiveness of disease-management programs for improving diabetes care: a meta-analysis. Can Med Assoc J 2011;183(2):E115-E27. doi: https://doi.org/10.1503/cmaj.091786 [published Online First: 2010/12/13].

96. Shojania KG, Ranji SR, McDonald KM, et al. Effects of quality improvement strategies for type 2 diabetes on glycemic control: a meta-regression analysis. Jama 2006;296(4):427–40. doi: https://doi.org/10.1001/jama.296.4.427 [published Online First: 2006/07/27].

97. Ampofo AG, Khan E, Ibitoye MB. Understanding the role of educational interventions on medication adherence in hypertension: A systematic review and meta-analysis. Heart Lung 2020 doi: https://doi.org/10.1016/j.hrtlng.2020.02.039 [published Online First: 2020/03/05].

98. Tong A, Manns B, Hemmelgarn B, et al. Establishing Core Outcome Domains in Hemodialysis: Report of the Standardized Outcomes in Nephrology-Hemodialysis (SONG-HD) Consensus Workshop. Am J Kidney Dis 2017;69(1):97–107. doi: https://doi.org/10.1053/j.ajkd.2016.05.022 [published Online First: 2016/08/09].

99. Kumar V, Roche C, Overman S, et al. What Is the Accuracy of Three Different Machine Learning Techniques to Predict Clinical Outcomes After Shoulder Arthroplasty? Clin Orthop Relat Res 2020 doi: https://doi.org/10.1097/corr.0000000000001263 [published Online First: 2020/04/26].

100. Harrison-Brown M, Scholes C, Ebrahimi M, et al. Predicting changes in the status of patient-reported outcome measures after Birmingham Hip Resurfacing: an observational cohort study with a median follow-up of ten years. Bone Joint J 2019;101-b(11):1431–37. doi: https://doi.org/10.1302/0301-620x.101b11.Bjj-2019-0663.R1 [published Online First: 2019/11/02].

101. Huber M, Kurz C, Leidl R. Predicting patient-reported outcomes following hip and knee replacement surgery using supervised machine learning. BMC Med Inform Decis Mak 2019;19(1):3. doi: https://doi.org/10.1186/s12911-018-0731-6 [published Online First: 2019/01/10].

102. Fontana MA, Lyman S, Sarker GK, et al. Can Machine Learning Algorithms Predict Which Patients Will Achieve Minimally Clinically Important Differences From Total Joint Arthroplasty? Clin Orthop Relat Res 2019;477(6):1267–79. doi: https://doi.org/10.1097/corr.0000000000000687 [published Online First: 2019/05/17].

103. Siccoli A, de Wispelaere MP, Schröder ML, et al. Machine learning-based preoperative predictive analytics for lumbar spinal stenosis. Neurosurg Focus 2019;46(5):E5. doi: https://doi.org/10.3171/2019.2.Focus18723 [published Online First: 2019/05/03].

104. Tennenhouse LG, Marrie RA, Bernstein CN, et al. Machine-learning models for depression and anxiety in individuals with immune-mediated inflammatory disease. J Psychosom Res 2020;134:110126. doi: https://doi.org/10.1016/j.jpsychores.2020.110126 [published Online First: 2020/05/11].

105. De Jong Y, Willik EM, Milders J, Voorend CGN, Morton RL, Dekker FW, Meuleman Y, Diepen M. A meta-review demonstrates improved reporting quality of qualitative reviews following the publication of COREQ- and ENTREQ-checklists, regardless of modest uptake. BMC Medical Research Methodology. https://doi.org/10.1186/s12874-021-01363-1.

Chapter 3

A META-REVIEW DEMONSTRATES IMPROVED REPORTING QUALITY OF QUALITATIVE REVIEWS FOLLOWING THE PUBLICATION OF COREQ- AND ENTREQ-CHECKLISTS, REGARDLESS OF MODEST UPTAKE

Ype de Jong
Esmee M. van der Willik
Jet Milders
Carlijn G. N. Voorend
Rachael L. Morton
Friedo W. Dekker
Yvette Meuleman
Merel van Diepen

BMC Medical Research Methodology. 2021 Sep 12;21(1):184
DOI: 10.1186/s12874-021-01363-1
PMID: 34511068
PMCID: PMC8436506.

To minimize the impact of printing on the environment, the supplementary material has been uploaded. For Chapter 3, the supplement is 102 pages, and can be downloaded using the QR code.

Abstract

Background: Reviews of qualitative studies allow for deeper understanding of concepts and findings beyond the single qualitative studies. Concerns on study reporting quality led to the publication of the COREQ-guidelines for qualitative studies in 2007, followed by the ENTREQ-guidelines for qualitative reviews in 2012. The aim of this meta-review is to: 1) investigate the uptake of the COREQ- and ENTREQ- checklists in qualitative reviews; and 2) compare the quality of reporting of the primary qualitative studies included within these reviews prior- and post COREQ-publication.

Methods: Reviews were searched on 02-Sept-2020 and categorized as (1) COREQ- or (2) ENTREQ-using, (3) using both, or (4) non-COREQ/ENTREQ. Proportions of usage were calculated over time. COREQ-scores of the primary studies included in these reviews were compared prior- and post COREQ-publication using T-test with Bonferroni correction.

Results: 1.695 qualitative reviews were included (222 COREQ, 369 ENTREQ, 62 both COREQ/ENTREQ and 1.042 non-COREQ/ENTREQ), spanning 12 years (2007–2019) demonstrating an exponential publication rate. The uptake of the ENTREQ in reviews is higher than the COREQ (respectively 28% and 17%), and increases over time. COREQ-scores could be extracted from 139 reviews (including 2.775 appraisals). Reporting quality improved following the COREQ-publication with 13 of the 32 signaling questions showing improvement; the average total score increased from 15.15 to 17.74 (p-value <0.001).

Conclusion: The number of qualitative reviews increased exponentially, but the uptake of the COREQ and ENTREQ was modest overall. Primary qualitative studies show a positive trend in reporting quality, which may have been facilitated by the publication of the COREQ.

3.1 INTRODUCTION

Qualitative studies allow for a deeper understanding of people's experiences, beliefs, attitudes or behaviors. These studies usually focus on *why* participants think or act in a certain way, using open ended data gathering methods such as interviews, focus groups or observations[1, 2]. They can be regarded as hypothesis generating research, and while research methods fundamentally differ when compared to quantitative research, they are not necessarily incompatible nor mutually exclusive. Both methods can complement each other, for example hypotheses that originated from qualitative research may be statistically tested in quantitative research, or findings from quantitative research can be explained by qualitative research[3, 4]. As in all fields of research, poorly designed, conducted or reported qualitative studies can lead to inappropriate findings[5].

In 2007, the COREQ (Consolidated criteria for reporting qualitative research) checklist was developed to assess the reporting quality of qualitative studies[6]. Realizing that, in contrast to most other research fields, no widely used comprehensive checklist, nor uniform and accepted requirements for publication of qualitative research existed, the authors aimed to "*… promote complete and transparent reporting among researchers and indirectly improve the rigor, comprehensiveness and credibility of interview and focus-group studies.*"[6] Items from 22 published checklists were compiled into a single 32-item checklist and grouped into three domains (*research team and reflexivity*, *study design* and *data analysis and reporting*), thus creating a comprehensive checklist covering the main aspects of qualitative research.

Though aimed at researchers conducting an interview- or focus group study, the COREQ also became frequently used in reviews on qualitative studies to assess the reporting quality of the included studies in the absence of a checklist specifically developed for this purpose. Qualitative reviews, a novel study design, aim to systematically synthesize the included qualitative studies instead of generating original data to achieve abstraction and transferability at a higher level beyond the included original studies[7, 8]. While in 2007, when the COREQ was published, the number of qualitative reviews was relatively limited, in 2012 this number had increased substantially. Thus, using a similar approach as the COREQ, in 2012 members from the same research team and international experts developed the ENTREQ (Enhancing transparency in reporting the synthesis of qualitative research) checklist, for reviews as opposed to original studies[9]. This 21-item checklist covers five domains (*introduction, methods and methodology, literature search and selection, appraisal,* and *synthesis of findings)* and aims to *"… develop a framework for reporting the synthesis of qualitative health research.*"[9]

Since the publication of both checklists, a large number of reviews of qualitative studies has been published on a wide array of topics. Though it has been argued that reporting checklists for qualitative research would not necessarily result in *better* research[10], and neither checklists were developed following the now accepted methods for developing reporting standards[11], both the COREQ and the ENTREQ are now included in the EQUATOR network[12], and are required by many clinical journals for submission; the high number of citations (respectively over 5.600 and 700 in Web of Science) indeed indicate usage. To this date however, no studies have been conducted to explore the uptake of the COREQ and the ENTREQ in reviews, or the effect on the reporting quality, which for guidelines in other research methods has been the case[13,14,15,16,17,18]. Therefore, the aim of this meta-review is twofold: 1) to investigate the uptake of the COREQ and ENTREQ checklists in reviews of primary qualitative studies, and 2) to compare the quality of reporting of the original qualitative studies included in these reviews prior- and post-publication of the COREQ.

3.2 METHODS

This meta-review was reported in line with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines[19].

3.2.1 Search strategy

Using similar searching methods as in previous studies, we developed three searches: the first search aimed to identify all qualitative reviews that cited the COREQ, the second aimed to identify all reviews that cited the ENTREQ; for these two searches, we used Web of Science and PubMed. Next, using terms encountered in these reviews, and building upon previous studies[20,21,22,23], we developed a comprehensive search method in PubMed to identify those reviews that did not specifically mention the COREQ or the ENTREQ. We then refined this broad search in an iterative process described in detail in the **Supplementary material chapter 3, section A**, and recoded the query to four other electronic databases: Cochrane library, Embase, Emcare and Web of Science. Searches were designed in collaboration with an experienced medical librarian and conducted on the 2nd of September 2020, including all articles since database inception (which differed per database). We then subtracted the results of the two other searches from this dataset. In the end, we thus obtained three databases: 1) studies citing the COREQ, 2) studies citing the ENTREQ, and 3) studies citing neither COREQ nor ENTREQ.

3.2.2 Eligibility methods

Studies were eligible for inclusion if they were 1) a review and 2) contained qualitative or mixed-methods research approaches. We created four datasets: reviews using the 1) COREQ, 2) ENTREQ, 3) both the COREQ and ENTREQ and 4) neither the COREQ or ENTREQ. To be included in the respective datasets, reviews using the COREQ were required to appraise their included studies with this checklist; those using the ENTREQ were required to mention adherence to it. Reviews were imported in Endnote (version 9.1) and duplicates were removed. One author (YdJ) screened the titles for obvious irrelevance. Two authors (YdJ and JM) independently selected studies for eligibility based on abstract and full-text; conflicts were resolved after discussion. The selection procedure is explained in more detail in the **Supplementary material chapter 3, section A**.

3.2.3 Data-extraction

Our study aimed to assess the uptake of the COREQ- and ENTREQ-checklists in reviews, but also to explore the effect of the COREQ on the reporting quality of original qualitative studies included in these reviews. For all reviews, we extracted the number of included qualitative studies, studies with mixed-method designs, and other designs (e.g. quantitative, reviews, etc.). For the first aim, we used the publication date of all the reviews from the meta-data of these reviews, rounded down to the month (i.e. MM/YYYY); if unavailable, we searched for the earliest publication date in online sources. For the second aim, we extracted the publication year (i.e. YYYY) and the COREQ scores of the original studies included in these reviews, as scored by the authors of these reviews (i.e. we did not rate the studies ourselves, but used the COREQ score as determined by the authors of these reviews, as illustrated in the **Supplementary material chapter 3, Figure S1**). Data were extracted on three levels based on availability of the data: the score at the level of signaling questions (reported or not reported; 0 or 1), the total score per domain (0–8 for domain 1, 0–15 for domain 2, and 0–9 for domain 3), and the overall total score (0–32), where applicable. If no extractable information (e.g. no review COREQ score, but only an average per domain) was available, the corresponding author of that study was contacted. Data extraction was conducted by YdJ, JM, EvdW, and CV; all experienced in qualitative research, and familiar with both the COREQ and ENTREQ checklists.

Figure 3.1 Schematic representation of inclusion periods used to assess the impact of the publication of the COREQ on the quality of reporting of the original qualitative studies (COREQ) included in qualitative reviews. For the COREQ, the COREQ score as assessed by the authors of the review was extracted and plotted over time using the publication date of that original qualitative study. All studies prior 2007 (so until and including 2006) were included, as were all those published after 2008 (so from and including 2009)

3.2.4 Statistical analysis

For the first aim, to investigate the uptake of the COREQ and the ENTREQ, we plotted the number of qualitative reviews using these checklists compared to those that did not use it over time, starting from the respective publication dates (i.e. 09–2007 and 11–2012). For the second aim, to assess whether the publication of the COREQ influenced the reporting quality of qualitative studies, we compared the average scores at the three levels (total score, domain scores and signaling questions) before publication of the COREQ (pre-COREQ: all studies before 2007) and after publication of the COREQ (post-COREQ: 2009–2019). Articles published in 2007 and 2008 were excluded, as the COREQ was published in September 2007 and this was regarded as a transition period, see **Figure 3.1**. We used this transition period to avoid inclusion of studies that used a preliminary version of the COREQ (which was presented at a congress prior to publication – personal communication with Prof. A. Tong), and also to exclude studies that were in the submission process at the time of the publication date. To visualize the trends of the total COREQ score per domain, we plotted the absolute score over time, using a LOESS curve with a 95% confidence interval, and a span of 0.5. Average scores, as opposed to median scores, were calculated as in similar prior studies[24, 25], as this allows comparison on the level of signaling questions, increase precision of the estimated effect, and, though fundamentally different than LOESS modelling, allows comparison to these curves more than median scores. To compare the average scores prior- and post publication, we used unpaired T-tests. As some COREQ scores were missing, analyses were performed on complete cases. A significance level of $p \leq 0.05$ was used, which was corrected for multiple testing using the Bonferroni approach. For the COREQ-analyses, we used a significance level of $p < 0.0014$ (0.05 divided by a total of 36 significance tests: 32 signaling questions, three domains and one for the total COREQ score). Analyses were performed in R, version 1.2.5001.

3.2.5 Sensitivity analyses

We conducted three sensitivity analyses, all related to the second aim. 1) An analysis where we compared the COREQ scores prior- and post-publication without the transition period. 2) An analysis after imputation of missing COREQ scores, since a substantial number of reviews presented an adapted or incomplete COREQ score, usually without explanation. We assumed these missing data to be missing at random (MAR) and conducted five-multiple imputations using the R-package MICE; estimates were pooled according to Rubin's rules. 3) An analysis of the effect of the inclusion of duplicate studies across reviews. Studies were considered a duplicate if the year of publication and name of the first author were identical. A detailed description of the sensitivity analyses is presented in the **Supplementary material chapter 3, section C.**

3.3 RESULTS

3.3.1 Characteristics of included studies

The three searches resulted in a total of 1.695 eligible reviews: 222 reviews used the COREQ for appraisal of their included studies, 369 used the ENTREQ, 62 reviews used both the COREQ and ENTREQ, and 1.042 used neither the COREQ or ENTREQ (**Figure 3.2**). These 1.695 reviews included a total of 49.281 studies (median 19 studies per review, IQR 12–32), most of which were qualitative (78%, 38.279; median 14 studies per review, IQR 8–26). The remaining studies were of mixed-methods (4%; 2.177 studies; median 2 studies per review, IQR 1–4) and other methodology (18%; 8.825 studies; median 11 studies per review, IQR 5–22). A summary of the included reviews is presented in **Table 3.1**; an overview of all included reviews is given in the **Supplementary material chapter 3, section D.**

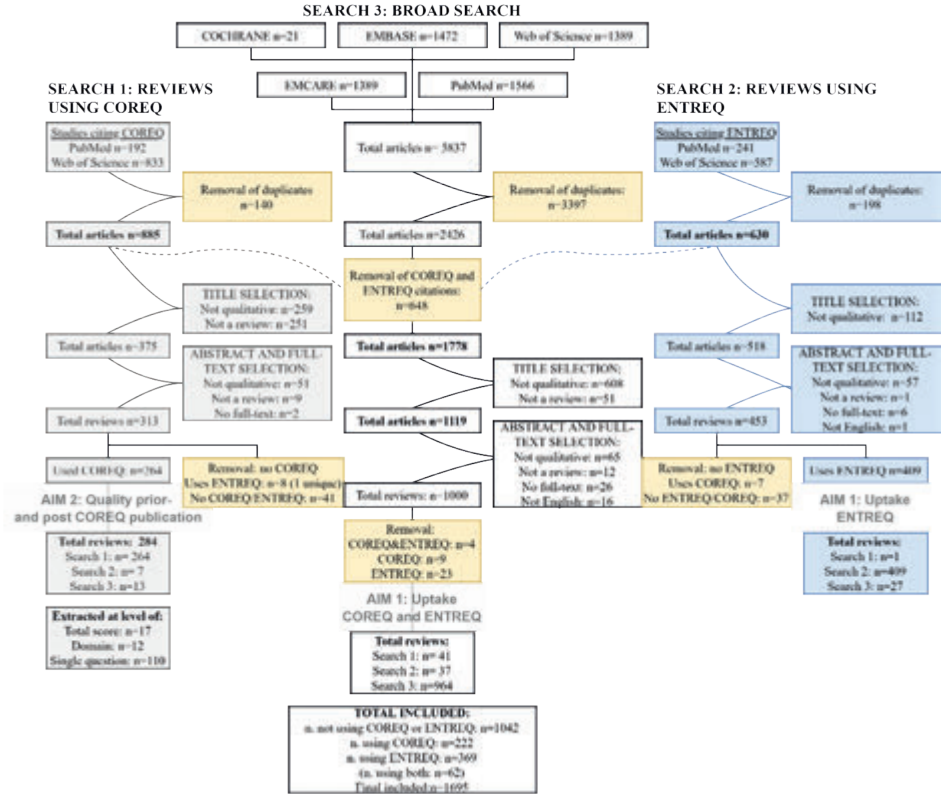


Figure 3.2 PRISMA flowchart of study inclusion. Three searches were conducted: the first search aimed to identify all reviews on qualitative research. The second and third searches were conducted in PubMed and Web of Science, and aimed to identify all reviews citing the COREQ and ENTREQ respectively. *The number of studies citing the COREQ or ENTREQ were subtracted from the total number of studies of search 1; since some studies cited both COREQ and ENTREQ, the total number of studies subtracted is less than the total numbers identified by the COREQ and ENTREQ searches

3.3.2 Characteristics of reviews using the COREQ

For the 282 reviews that used the COREQ (i.e. 222 reviews using the COREQ alone; 62 using both COREQ and ENTREQ), most reviews presented their appraisal results in a table (n=193; 68%), or textual only (n=37, 13%), or a bar chart (n= 3, 1%). A large number of reviews appraised their included studies with the COREQ, but did not present the results (49 reviews; 17%). A total of 139 (49%) of the 282 reviews presented extractable data from individual studies, which was used to explore the trends in COREQ scores over time. Of these 139 reviews, data were presented at the level of signaling questions for 110 (79%), domains for 12 (9%) and total score for 17 (12%) of the reviews. In total, 2.775 COREQ appraisals of qualitative studies were extracted: 2.448 at the level of signaling questions, 200 at domain score, and 127 at overall total score. In more than half of the

reviews, the COREQ checklist was adapted for study purposes (e.g. item exclusion) or COREQ-scores were incompletely reported: 47 out of the 110 reviews that reported at the level of signaling questions scored at least one of their included studies on all 32 signaling questions. The median completeness of the 32 COREQ-items was 25 (IQR 23–32; range 1–32), for the completeness of the individual signaling questions, see **Table 3.2**. As we used only the complete scores for our analyses (i.e. a complete case analysis), the number of appraisals included in the analysis for COREQ domains 1 to 3 was 1.036, 1.117, 1.086 respectively, and 831 appraisals for the overall total COREQ score.

	Total	**COREQ**	**ENTREQ**	**Both COREQ/ ENTREQ**	**Non-COREQ/ ENTREQ**
Total reviews (% of total)	1.695	222 (13%)	369 (22%)	62 (4%)	1.042 (61%)
Characteristics					
Included studies	49.281	6.069	9.715	2.042	31.455
Median (IQR)	19 (12 - 32)	20.50 (13 - 31.75)	18 (11 - 32)	28 (14.25 - 42.75)	19 (12 - 31)
Qualitative (% of total)	38.279 (78%)	3.527 (58%)	8.282 (85%)	1.915 (94%)	24.555 (78%)
Median (IQR)	14 (8 - 26)	10 (5 - 20)	14 (9 - 28.50)	26 (13 - 39.75)	15 (9 - 25)
Mixed methods (% of total)	2.177 (4%)	349 (6%)	453 (5%)	22 (1%)	1.353 (4%)
Median (IQR)	2 (1 - 4)	2 (2 - 5)	2 (1 - 4)	1.5 (1 - 3.50)	2 (1 - 4)
Other (% of total)*	8.825 (18%)	2.193 (36%)	980 (10%)	105 (5%)	5.547 (18%)
Median (IQR)	11 (5 - 22)	12 (7 - 21.25)	11 (5 - 19)	12 (8.50 - 29.75)	11 (5 - 22)

Table 3.1 Summary of the 1.695 included qualitative reviews, grouped as COREQ- or ENTREQ using, using both checklists, or using neither checklist. An overview of each included review is presented in the supplement, section D. *Other study design includes all studies that are neither qualitative or mixed methods (e.g. quantitative, reviews, etc.)

3.3.3 First aim: trends over time: uptake of COREQ and ENTREQ over time

The total number of reviews on qualitative studies increased exponentially over time (**Figure 3.3A**). Until the publication of the COREQ in September 2007, only 31 reviews were identified; this number increased to 141 at the publication of the ENTREQ in November 2012. Of the total of 1.664 reviews published since the COREQ publication, 284 (17%) used the COREQ to assess the reporting quality of their included studies, this proportion remaining stable over time (**Figure 3.3B and C**). For the ENTREQ, 431 reviews (28%) used this checklist out of the 1.554 reviews published since its publication, with this proportion increasing over time (**Figure 3.3B and D**).

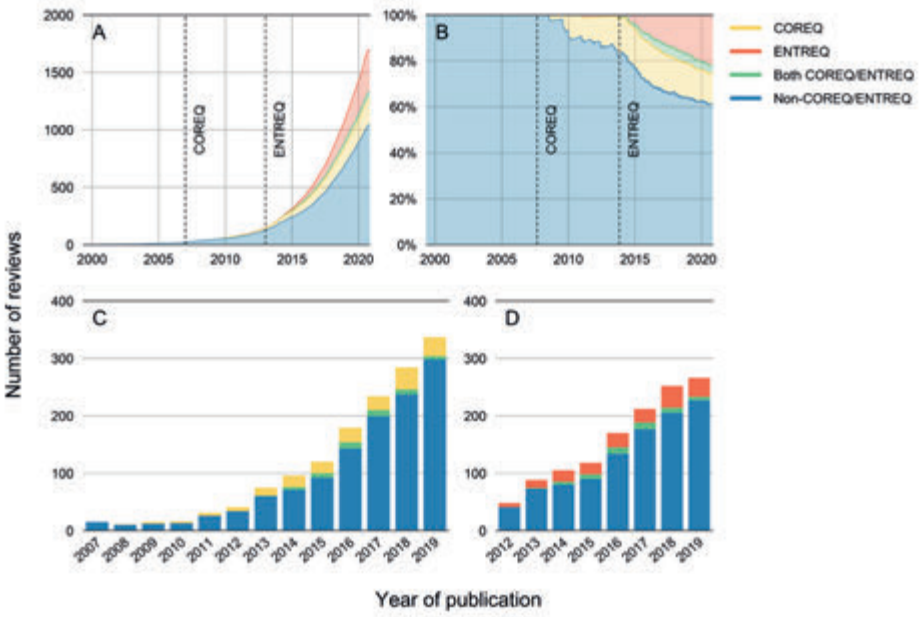


Figure 3.3 Uptake of the COREQ and the ENTREQ. 3A Stacked chart of qualitative reviews over time. 3B Percent stacked chart, showing the (cumulative) proportion of COREQ, ENTREQ and non-COREQ/ENTREQ reviews over time. 3C Absolute number of COREQ versus non-COREQ (including ENTREQ and non-COREQ/ENTREQ reviews) stratified per year since the publication of the COREQ in 2007. 3D Absolute number of ENTREQ versus non-ENTREQ (including COREQ and non-COREQ/ENTREQ reviews) stratified per year since the publication of the ENTREQ in 2012

3.3.4 Second aim: reporting quality prior- and post-publication of the COREQ

Of the 2.775 studies that were appraised with the COREQ, a total of 1.045 (39%) were published before 2007 and 1.415 (51%) after 2008; we thus excluded 315 (11%) studies for this analysis. The total COREQ score increased from 15.51 (SE 0.31) to 17.74 (SE 0.20, p-value <0.001). The average scores per domain prior- and post-publication all increased: *research team and reflexivity:* 2.57, SE 0.12 before 2007 and 2.86, SE 0.08 after 2008 (difference 0.29, p-value 0.048), *study design*: 7.97, SE 0.15 before 2007 and 8.51, SE 0.10 after 2008 (difference 0.55, p-value 0.007), and *data analysis and reporting*: 5.42, SE 0.10 before 2007 and 6.20, SE 0.07 after 2008 (difference 0.78, p-value <0.001). After Bonferroni correction, 13 out of the 32 signaling questions showed improvement. An overview of the average scores per signaling questions both prior- and post-publication of the COREQ is presented in **Table 3.2**, the positive trendline for each of the three domains is visualized in **Figure 3.4**.

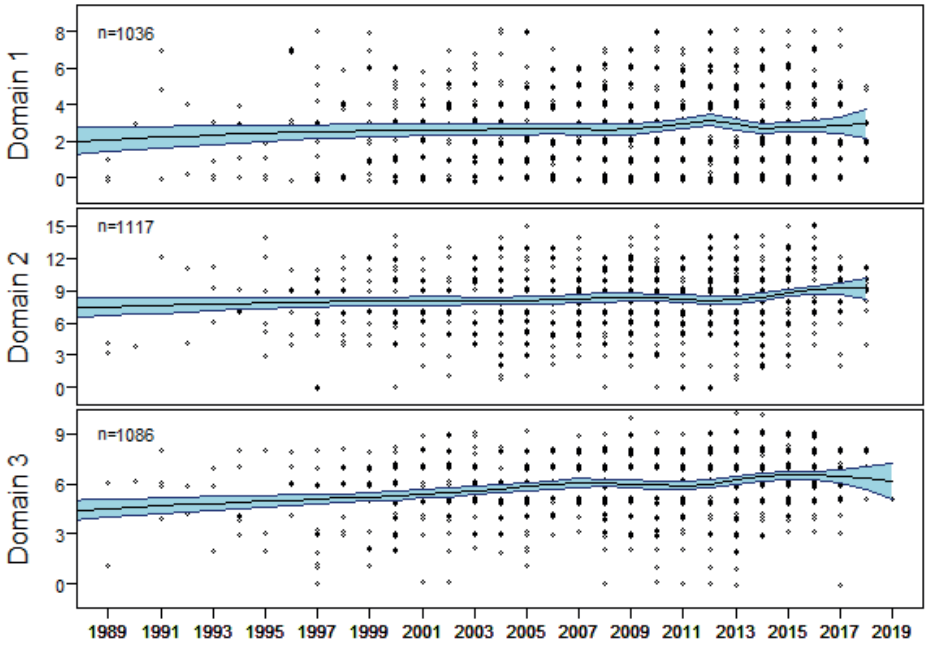


Figure 3.4 Trends for the three domains of the COREQ (domain 1: *research team and reflexivity;* domain 2: *study design* and domain 3: *data analysis and reporting)*, plotted over time, with a smoothed LOESS curve and 95% confidence interval (light blue). Y-axis differs per domain, as the number of signaling questions per domain is different (ranges per domain: 0–8 for domain 1, 0–15 for domain 2, and 0–9 for domain 3). For clarity, data points are jittered on the y-axis, by adding a Gaussian error with a standard deviation of 0.1

3.3.5 Sensitivity analyses

When comparing the COREQ without the transition period, the improvement was less pronounced with 11 out of the 32 signaling questions showing changes after Bonferroni correction (one negative, the others positive; see **Supplementary material chapter 3, Table S3**). For the second sensitivity analysis, we imputed the missing data assuming MAR. The results were similar, with 11 signaling questions showing a positive change (**Supplementary material chapter 3, Table S4**). Of the 2.775 studies included for the second aim, there were 185 (7%) studies included more than once (142 included two times, 31 included three times, and 12 included four or more times), resulting in a total duplicate count of 430. The results were similar to the main analysis, with 14 signaling questions showing a positive change (**Supplementary material chapter 3, Table S5**).

	TOTAL		PRE-COREQ (<2007)				POST-COREQ (>2008)				
	n	*complete*	*Score*	*n*	*Score*	*SE*	*n*	*Score*	*SE*	*Difference*	*P*
DOMAIN I: RESEARCH TEAM AND REFLEXIVITY	**1037**		**2.76**	**309**	**2.57**	**0.12**	**621**	**2.86**	**0.08**	**0.29**	**0.048**
Interviewer/facilitator	2218	91%	0.56	771	0.51	0.02	1187	0.58	0.01	0.08	0.001
Credentials	1227	50%	0.42	379	0.41	0.03	708	0.42	0.02	0.01	0.788
Occupation	2092	85%	0.43	730	0.43	0.02	1121	0.41	0.01	-0.02	0.372
Gender	1300	53%	0.43	460	0.38	0.02	685	0.45	0.02	0.07	0.026
Experience and training	2262	92%	0.25	799	0.20	0.01	1200	0.26	0.01	0.06	0.001
Relationship established	2193	90%	0.18	757	0.18	0.01	1181	0.18	0.01	0.00	0.836
Participant knowledge of the interviewer	1022	42%	0.16	308	0.15	0.02	603	0.16	0.02	0.01	0.581
Interviewer characteristics	922	38%	0.20	236	0.16	0.02	585	0.23	0.02	0.07	0.020
DOMAIN II: STUDY DESIGN	**1117**		**8.31**	**351**	**7.97**	**0.15**	**645**	**8.51**	**0.10**	**0.55**	**0.007**
Methodological orientation and Theory	1206	49%	0.72	391	0.70	0.02	662	0.74	0.02	0.03	0.258
Sampling	2337	95%	0.77	815	0.73	0.02	1247	0.79	0.01	0.06	0.003
Method of approach	2241	92%	0.71	785	0.66	0.02	1195	0.74	0.01	0.08	0.000
Sample size	2384	97%	0.95	838	0.94	0.01	1259	0.96	0.01	0.02	0.023
Non-participation	2229	91%	0.42	769	0.39	0.02	1203	0.45	0.01	0.06	0.008
Setting of data collection	2377	97%	0.67	833	0.65	0.02	1258	0.67	0.01	0.02	0.339
Presence of nonparticipants	2237	91%	0.24	772	0.20	0.01	1207	0.26	0.01	0.06	0.001
Description of sample	2394	98%	0.88	848	0.83	0.01	1259	0.90	0.01	0.07	0.000
Interview guide	2329	95%	0.72	817	0.66	0.02	1234	0.77	0.01	0.12	0.000
Repeat interviews	2195	90%	0.30	769	0.34	0.02	1177	0.28	0.01	-0.06	0.005
Audio/visual recording	2262	92%	0.79	782	0.70	0.02	1219	0.83	0.01	0.13	0.000
Field notes	2314	95%	0.33	814	0.29	0.02	1224	0.34	0.01	0.05	0.022
Duration	2288	93%	0.61	793	0.57	0.02	1223	0.63	0.01	0.06	0.007
Data saturation	2157	88%	0.29	753	0.20	0.01	1147	0.35	0.01	0.16	0.000
Transcripts returned	1450	59%	0.13	494	0.11	0.01	789	0.15	0.01	0.04	0.028
DOMAIN III: ANALYSIS AND FINDINGS	**1086**		**5.94**	**339**	**5.42**	**0.10**	**625**	**6.20**	**0.07**	**0.78**	**0.000**
Number of data coders	2304	94%	0.60	785	0.49	0.02	1248	0.66	0.01	0.17	0.000
Description of the coding tree	1333	54%	0.43	474	0.39	0.02	703	0.47	0.02	0.08	0.004
Derivation of themes	2298	94%	0.80	800	0.74	0.02	1222	0.84	0.01	0.10	0.000
Software	2312	94%	0.39	801	0.28	0.02	1234	0.45	0.01	0.17	0.000
Participant checking	2259	92%	0.19	775	0.19	0.01	1225	0.19	0.01	0.01	0.752
Quotations presented	2384	97%	0.89	839	0.86	0.01	1260	0.92	0.01	0.06	0.000
Data and findings consistent	2187	89%	0.77	749	0.72	0.02	1186	0.79	0.01	0.08	0.000
Clarity of major themes	1085	44%	0.94	335	0.90	0.02	611	0.95	0.01	0.05	0.009
Clarity of minor themes	993	41%	0.65	309	0.62	0.03	565	0.66	0.02	0.04	0.303
TOTAL COREQ SCORE	**831**		**16.99**	**232**	**15.51**	**0.31**	**509**	**17.74**	**0.20**	**2.23**	**0.000**

Table 3.2 A complete-case comparison is made between those studies published prior to 2007 and those published after 2008. Because of this time-window, 315 studies were excluded for this analysis. Differences in mean scores were calculated by unpaired T-tests; significance ($p < 0.05$) is indicated by an asterisk (*); significance after Bonferroni correction (36 significance tests: 32 signaling questions, 3 domains, 1 total score, hence 0.05/36, $p \leq 0.0014$) is indicated by two asterisks (**). % complete denotes the completeness of reporting for that specific signaling question. Ranges per domain: 0–8 for domain 1, 0–15 for domain 2, and 0–9 for domain 3

3.4 DISCUSSION

In this meta-review, we explored the uptake of the COREQ- and ENTREQ-checklists in qualitative reviews, and compared the reporting quality of original qualitive studies prior- and post COREQ publication. Though reviews of qualitative research are a novel methodology to achieve abstraction beyond the original qualitative studies, we demonstrated an exponential publication trend over the past twenty years. By including 1.695 reviews, that in turn included 49.281 studies, we were able to present an in-depth overview of current qualitative research – both at the level of reviews, as well as the level of individual studies included within these reviews. Answering the first research question, we found that the COREQ, published in 2007 to score the quality of reporting of original qualitive studies, was used in 17% of the reviews to appraise the reporting quality of their included studies. The ENTREQ, published in 2012 specifically for systematic reviews, showed a better uptake with 28% of the reviews using the checklist. Finally, using the COREQ-scores of 2.775 studies within these reviews, we demonstrated a positive trend in reporting quality since the publication of the COREQ, with 13 out of the 32 signaling questions showing improvement.

The uptake of the COREQ in qualitative reviews may be explained by the original aim of the COREQ, namely to improve quality of reporting in original interview- or focus-group studies[6]. In the absence of a comprehensive checklist for reporting the quality of qualitative reviews, the usage of the COREQ to appraise the reporting quality of studies within reviews may have followed naturally with the increasing numbers of qualitative reviews since its publication. The ENTREQ, specifically designed for reviews, showed a higher uptake[9]. Yet, appraising qualitative studies remains a debated topic. While some argue that adhering to checklists improves transparency and validity of findings, others feel endorsement as a limitation, arguing that a 'one size fits all' -set of criteria cannot encompass the broadness of qualitative research as a whole[5, 26,27,28,29]. In our study, this unresolved debate is clearly illustrated by the large number of reviews that adapted the COREQ for their purposes: more than half of the studies assessed their included studies with a selection of COREQ-items, or combined it with other checklists, both designed for reporting- or overall quality assessment, such as the CASP[30], QualSys[31], GRADE-CREQual[32], MMAT[33], amongst others. The incomplete reporting, or the limited uptake of the COREQ and ENTREQ is not unique for qualitative research. For example, impact-studies on guidelines used for quantitative reviews[19], clinical trials[13, 34], observational studies[15, 16], prediction- or prognostic studies[14, 17], show that, even with endorsement of journals, the completeness of reporting remains suboptimal although for some, reporting quality improved.

By extracting the COREQ-scores of 2.775 appraisals included in these reviews, we were able to observe changes in the quality of reporting over time. On average, the total score, one of the three domains, and nearly half of the 32 signaling questions showed improvement when comparing studies published prior- versus post-publication of the COREQ. Though causal inferences cannot be made, this improvement, especially viewed in combination with the exponential trend of qualitative review publications, reflects the maturation and increasing acceptance of qualitative research. Although the overall quality of reporting improved, the scores of some items remained remarkably low: 16 out of the 32 signaling questions scored lower than an average score of 0.5. For example, in the first domain ("*research team and reflexivity*"), the items "*experience and training*", "*relationship established*" and "*participant knowledge of the interviewer*" were reported poorly and did not improve markedly, with an average score of 0.25, 0.18 and 0.16, meaning that only 25, 18 and 16% of the articles reported these items, respectively. For the second domain ("*study design*"), most items were reported better than in the first domain, and improvements were even stronger. Nearly all items improved, and almost half remained significant after Bonferroni correction for multiple testing. The third domain ("*analysis and findings*") showed good reporting on nearly all items, except for "*software*" and "*participant checking*", though the first showed the largest improvement of all 32 items of the COREQ. These findings are in line with the two other studies that graded qualitative studies for the same purpose: Al-Moghrabi et al graded 100 qualitative studies, and demonstrated poor quality of reporting for most signaling questions[31]. In the second study, Godinho et al confirms this poor completeness of reporting in 246 Indian qualitative studies[24, 25]. When plotting the results over time, completeness of reporting remained modest, but increased over time, possibly facilitated by the publication of the COREQ and subsequent endorsement of journals[30].

The strengths of this study are the large sample size and comprehensive search methods. We conducted our study on reviews of qualitative studies (i.e. a meta-review). This method allowed for exploration of checklist usage in the same study type, namely reviews. Furthermore, the original qualitative studies included in these reviews are independently assessed for reporting quality by the authors of these reviews, assuring independent quality assessment and allowing for a large number of study appraisals to be included. We aimed to include as many studies as possible, in order to present a comprehensive overview of all qualitative reviews. However, because of this large sample size, we did not perform complete cross-checking at two levels: title selection and data-extraction. We did cross-check the abstract- and full-texts for inclusion, showing excellent agreement (Cohen's kappa coefficient for inter-rater reliability of 0.86 and 1.00 respectively). Data-extraction

was cross-checked for 10 reviews, showing no errors. Furthermore, nearly all COREQ-studies could be extracted directly by recoding the COREQ-tables to our format, instead of typing the scores in our data-system, thus reducing the risk of errors. Next, though misclassification of study type could be a more serious issue (e.g. misclassify a qualitative study design as mixed methods), all authors used the same methodology to classify the study types, as detailed in the supplement. Another limitation related to the COREQ-score is selection bias: studies of higher quality may have been easier to find in database-searches than those that are of lower quality (e.g. because of the use of identifiable terms as 'thematic synthesis' or 'grounded theory'), possibly resulting in overestimation of the average COREQ scores. Furthermore, some review authors might have excluded studies based on their COREQ-score, which will result in an overestimation of the COREQ scores. Since the publication of the COREQ and ENTREQ, various new checklists have been published, both for appraising the reporting- and the overall study quality (e.g. the CASP in 2013[30], the SRQR checklist in 2014, the eMERGe in 2019), underlining the developments in this research field since these guidelines. The use of these guidelines might partly explain the limited uptake of the COREQ and the ENTREQ, however we believe this to be to a limited extent since most reviews that did not use the COREQ or ENTREQ did not use any other checklist. Another explanation of the limited uptake may be improved retrievability of the post-COREQ and ENTREQ studies: including terms as 'adhering to', 'appraising', or naming these checklists likely increased the likelihood of inclusion in our review, compared to studies published prior these guidelines. Because of this, we based our search on previous studies[22, 23], designed our queries together with an experienced medical librarian, and conducted iterative search methods, and we thus believe this effect to be minimal. Lastly, it cannot be inferred that differences prior- and post-publication of the COREQ and ENTREQ are causally related to the publication of these checklists.

3.4.1 Implications and conclusion

Our study highlights several points that may further improve the quality of reporting. First, surprisingly, almost a fifth of the reviews that used the COREQ did not present the results of their quality appraisal. Given that four out of the 21 ENTREQ-items, but also four of the 27 PRISMA-items concern study appraisal, at least reporting appraisal results should be the minimum. Ideally however, to facilitate meta-reviews of this kind, and to increase transparency and reproducibility, reporting appraisal results per individual study at the level of signaling questions is essential. Next, though we did not explore the characteristics of the authors of our included reviews, it can reasonably be assumed that the exponential publication trend may be explained by an increasing number of unique

authors. Whether or not articles should be scored instead of appraised in a descriptive way remains open for discussion. However, the use of these checklists might be beneficial for new or inexperienced authors designing a qualitative study: checklists may guide those unfamiliar with qualitative research with hints and directions to avoid commonly made mistakes[5, 10, 27, 35]. The same holds true for reviewers assessing a qualitative review for publication, particularly if the reviewer has content expertise but not methodological expertise. A final implication concerns the poor reporting of several signaling questions of the COREQ. Whether or not these items are intentionally or unintentionally underreported, our study clearly points towards items that might either actually improve qualitative research if reported, or be left out from the checklist in a possible later or updated version. By providing this information on a large number of qualitative studies, our study might thus facilitate the ongoing discussions by providing factual data on both the use of checklists, and the completeness of reporting.

3.5 REFERENCES

1. Kuper A, Reeves S, Levinson W. An introduction to reading and appraising qualitative research. BMJ. 2008;337:a288. https://doi.org/10.1136/bmj.a288 [published Online First: 2008/08/09].
2. Giacomini MK, Cook DJ. Users' guides to the medical literature: XXIII. Qualitative research in health care B. What are the results and how do they help me care for my patients? Evidence-Based Medicine Working Group. JAMA. 2000;284(4):478–82. https://doi.org/10.1001/jama.284.4.478 [published Online First: 2000/07/25].
3. O'Cathain A, Thomas KJ, Drabble SJ, et al. What can qualitative research do for randomised controlled trials? A systematic mapping review. BMJ Open. 2013;3(6). https://doi.org/10.1136/bmjopen-2013-002889 [published Online First: 2013/06/26].
4. Lewin S, Glenton C, Oxman AD. Use of qualitative methods alongside randomised controlled trials of complex healthcare interventions: methodological study. BMJ. 2009;339:b3496. https://doi.org/10.1136/bmj.b3496 [published Online First: 2009/09/12].
5. Reynolds J, Kizito J, Ezumah N, et al. Quality assurance of qualitative research: a review of the discourse. Health Res Policy Syst. 2011;9:43. https://doi.org/10.1186/1478-4505-9-43 [published Online First: 2011/12/21].
6. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care. 2007;19(6):349–57. https://doi.org/10.1093/intqhc/mzm042 [published Online First: 2007/09/18].
7. Dixon-Woods M, Agarwal S, Jones D, et al. Synthesising qualitative and quantitative evidence: a review of possible methods. J Health Serv Res Policy. 2005;10(1):45–53. https://doi.org/10.1177/135581960501000110 [published Online First: 2005/01/26].
8. Butler A, Hall H, Copnell B. A Guide to Writing a Qualitative Systematic Review Protocol to Enhance Evidence-Based Practice in Nursing and Health Care. Worldviews Evid Based Nurs. 2016;13(3):241–9. https://doi.org/10.1111/wvn.12134 [published Online First: 2016/01/21].
9. Tong A, Flemming K, McInnes E, et al. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. BMC Med Res Methodol. 2012;12:181. https://doi.org/10.1186/1471-2288-12-181 [published Online First: 2012/11/29].
10. Hannes K, Heyvaert M, Slegers K, et al. Exploring the potential for a consolidated standard for reporting guidelines for qualitative research: an argument Delphi approach. Int J Qual Methods. 2015;14(4):1609406915611528. https://doi.org/10.1177/1609406915611528.
11. Moher D, Schulz KF, Simera I, et al. Guidance for developers of Health Research reporting guidelines. PLoS Med. 2010;7(2):e1000217. https://doi.org/10.1371/journal.pmed.1000217.
12. EQUATOR Network: Enhancing the QUAlity and Transparency Of health Research [Available from: https://www.equator-network.org/ accessed 11-11-2020.
13. Moher D, Jones A, Lepage L. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. JAMA. 2001;285(15):1992–5. https://doi.org/10.1001/jama.285.15.1992 [published Online First: 2001/04/20].
14. Zamanipoor Najafabadi AH, Ramspek CL, Dekker FW, et al. TRIPOD statement: a preliminary pre-post analysis of reporting and methods of prediction models. BMJ Open. 2020;10(9):e041537–e37. https://doi.org/10.1136/bmjopen-2020-041537.
15. Bastuji-Garin S, Sbidian E, Gaudy-Marqueste C, et al. Impact of STROBE statement publication on quality of observational study reporting: interrupted time series versus before-after analysis. PLoS One. 2013;8(8):e64733. https://doi.org/10.1371/journal.pone.0064733 [published Online First: 2013/08/31].

16. Poorolajal J, Cheraghi Z, Irani AD, et al. Quality of Cohort Studies Reporting Post the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement. Epidemiol Health. 2011;33:e2011005. https://doi.org/10.4178/epih/e2011005 [published Online First: 2011/07/01].

17. Sekula P, Mallett S, Altman DG, et al. Did the reporting of prognostic studies of tumour markers improve since the introduction of REMARK guideline? A comparison of reporting in published articles. PLoS One. 2017;12(6):e0178531. https://doi.org/10.1371/journal.pone.0178531 [published Online First: 2017/06/15].

18. Smidt N, Rutjes AW, van der Windt DA, et al. The quality of diagnostic accuracy studies since the STARD statement: has it improved? Neurology. 2006;67(5):792–7. https://doi.org/10.1212/01.wnl.0000238386.41398.30 [published Online First: 2006/09/13].

19. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6(7):e1000097. https://doi.org/10.1371/journal.pmed.1000097 [published Online First: 2009/07/22].

20. Ype, Jong Chava L., Ramspek Carmine, Zoccali Kitty J., Jager Friedo W., Dekker Merel, Diepen Appraising prediction research: a guide and meta-review on bias and applicability assessment using the Prediction model Risk Of Bias ASsessment Tool (PROBAST). Nephrology. https://doi.org/10.1111/nep.13913.

21. Ype Jong, Esmee M. Willik, Jet Milders, Yvette Meuleman, Rachael L. Morton, Friedo W. Dekker, Merel Diepen Person centred care provision and care planning in chronic kidney disease: which outcomes matter? A systematic review and thematic synthesis of qualitative studies. BMC Nephrology. https://doi.org/10.1186/s12882-021-02489-6.

22. Walters LA, Wilczynski NL, Haynes RB. Developing optimal search strategies for retrieving clinically relevant qualitative studies in EMBASE. Qual Health Res. 2006;16(1):162–8. https://doi.org/10.1177/1049732305284027 [published Online First: 2005/12/01].

23. Barroso J, Gollop CJ, Sandelowski M, et al. The challenges of searching for and retrieving qualitative studies. West J Nurs Res. 2003;25(2):153–78. https://doi.org/10.1177/0193945902250034 [published Online First: 2003/04/02].

24. Godinho MA, Gudi N, Milkowska M, et al. Completeness of reporting in Indian qualitative public health research: a systematic review of 20 years of literature. J Public Health. 2019;41(2):405–11. https://doi.org/10.1093/pubmed/fdy122.

25. Al-Moghrabi D, Tsichlaki A, Alkadi S, et al. How well are dental qualitative studies involving interviews and focus groups reported? J Dent. 2019;84:44–8. https://doi.org/10.1016/j.jdent.2019.03.001 [published Online First: 2019/03/14].

26. Buus N, Agdal R. Can the use of reporting guidelines in peer-review damage the quality and contribution of qualitative health care research? Int J Nurs Stud. 2013;50(10):1289–91. https://doi.org/10.1016/j.ijnurstu.2013.02.012 [published Online First: 2013/03/20].

27. Barbour RS. Checklists for improving rigour in qualitative research: a case of the tail wagging the dog? BMJ. 2001;322(7294):1115–7. https://doi.org/10.1136/bmj.322.7294.1115 [published Online First: 2001/05/05].

28. Dixon-Woods M, Shaw RL, Agarwal S, et al. The problem of appraising qualitative research. Qual Saf Health Care. 2004;13(3):223–5. https://doi.org/10.1136/qhc.13.3.223 [published Online First: 2004/06/04].

29. Kuper A, Lingard L, Levinson W. Critically appraising qualitative research. BMJ. 2008;337:a1035. https://doi.org/10.1136/bmj.a1035 [published Online First: 2008/08/09].

30. Critical Appraisal Skills Programme (2018) CASP Qualitative checklist [Available from: https://casp-uk.net/wp-content/uploads/2018/03/CASP-Qualitative-Checklist-2018_fillable_form.pdf accessed accessed 14-10-2020.

31. Kmet L, Lee R. Standard quality assessment criteria for evaluating primary research papers from a variety of FieldsAHFMRHTA Initiative20040213. HTA Initiative. 2004;2. https://www.ihe.ca/publications/standard-quality-assessment-criteria-for-evaluating-primary-research-papers-froma-variety-of-fields.

32. Lewin S, Glenton C, Munthe-Kaas H, et al. Using qualitative evidence in decision making for health and social interventions: an approach to assess confidence in findings from qualitative evidence syntheses (GRADE-CERQual). PLoS Med. 2015;12(10):e1001895. https://doi.org/10.1371/journal.pmed.1001895 [published Online First: 2015/10/28].

33. Hong QN, Fàbregues S, Bartlett G, et al. The mixed methods appraisal tool (MMAT) version 2018 for information professionals and researchers. Educ Inf. 2018;34:1–7. https://doi.org/10.3233/EFI-180221.

34. Turner L, Shamseer L, Altman DG, et al. Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. Syst Rev. 2012;1:60. https://doi.org/10.1186/2046-4053-1-60 [published Online First: 2012/12/01].

35. Vandenbroucke JP. STREGA, STROBE, STARD, SQUIRE, MOOSE, PRISMA, GNOSIS, TREND, ORION, COREQ, QUOROM, REMARK... and CONSORT: for whom does the guideline toll? J Clin Epidemiol. 2009;62(6):594–6. https://doi.org/10.1016/j.jclinepi.2008.12.003 [published Online First: 2009/02/03].

PART II

Prediction of ischemic stroke in CKD and dialysis: methodological quality of current models and clinical implications

Chapter 4

A SYSTEMATIC REVIEW AND EXTERNAL VALIDATION OF STROKE PREDICTION MODELS DEMONSTRATES POOR PERFORMANCE IN DIALYSIS PATIENTS

Ype de Jong
Chava L. Ramspek
Vera H.W. van der Endt
Maarten B. Rookmaaker
Peter J. Blankestijn
Robin W.M. Vernooij
Marianne C. Verhaar
Willem Jan W. Bos
Friedo W. Dekker
Gurbey Ocak
Merel van Diepen

Journal of Clinical Epidemiology 2020 Jul;123:69-79.
DOI: 10.1016/j.jclinepi.2020.03.015.
PMID: 32240769.

To minimize the impact of printing on the environment, the supplementary material has been uploaded. For Chapter 4, the supplement is 36 pages, and can be downloaded using the QR code.

Abstract

Objectives: The objective of this study was to systematically review and externally assess the predictive performance of models for ischemic stroke in incident dialysis patients.

Study design and setting: Two reviewers systematically searched and selected ischemic stroke models. Risk of bias was assessed with the PROBAST. Predictive performance was evaluated within The Netherlands Cooperative Study on the Adequacy of Dialysis (NECOSAD), a large prospective multicenter cohort of incident dialysis patients. For discrimination, c-statistics were calculated; calibration was assessed by plotting predicted and observed probabilities for stroke, and calibration-in-the-large.

Results: Seventy-seven prediction models for stroke were identified, of which 15 were validated. Risk of bias was high, with all of these models scoring high risk in one or more domains. In NECOSAD, of the 1,955 patients, 127 (6.5%) suffered an ischemic stroke during the follow-up of 2.5 years. Compared with the original studies, most models performed worse with all models showing poor calibration and discriminative abilities (c-statistics ranging from 0.49 to 0.66). The Framingham showed reasonable calibration; however, with a c-statistic of 0.57 (95% CI 0.50-0.63), the discrimination was poor.

Conclusion: This external validation demonstrates the weak predictive performance of ischemic stroke models in incident dialysis patients. Instead of using these models in this fragile population, either existing models should be updated, or novel models should be developed and validated.

4.1 INTRODUCTION

Stroke is a leading cause of morbidity and mortality worldwide. While mortality rates are declining, incidence rates and disease burden have increased over the years[1]. Stroke rates increase with declining renal function and reach a fivefold to tenfold increase in end-stage renal disease patients on dialysis compared with the general population[2-5]. Furthermore, the prognosis in patients on dialysis suffering from a stroke is generally poor: hemodialysis patients have a 3-fold higher risk of death after acute stroke compared with non-dialysis populations[5,6].

Identification of those dialysis patients at increased risk for stroke is thus of major importance. Prediction models that assess the risk of stroke, such as the commonly used CHA_2DS_2-$VASc_2$[7] and $CHADS_2$[8] have been developed and validated to efficiently allocate individualized anticoagulation therapy. External validation, a step which is essential before implementation of prediction models, shows reasonable predictive performance in independent cohorts with similar characteristics as the development cohorts of these models. However, dialysis patients were not included in the development of these models, and predictive performance within this high-risk population is largely unknown: only the CHA_2DS_2-$VASc_2$ and $CHADS_2$ have been externally validated. One study reported modest discrimination in a prevalent cohort of dialysis patients[9], and another found good predictive performance in a small cohort of dialysis patients with atrial fibrillation[10]. However, many more prediction models exists and are commonly used in clinical practice, despite the uncertainties regarding predictive performance in this fragile population. In addition, as weighing the benefits of anticoagulation versus the increased risk of bleeding is essential, we have previously conducted an external validation of bleeding risk models which showed poor predictive performance in incident dialysis patients[11]. To further contribute to the ongoing discussion on stroke management in dialysis patients, the aim of the present study is to provide a systematic review and independent external validation of stroke risk models in incident dialysis patients.

4.2 METHODS

4.2.1 Systematic review

The current review was designed to identify prediction models that assess the risk of ischemic stroke in any population. The PRISMA[12], TRIPOD[13], and CHARMS[14] guidelines were followed to ensure transparent reporting.

4.2.1.1 Study selection

Studies were included if they met the following predefined selection criteria: (1) The study developed a multivariable prognostic prediction model, with a prediction research question as aim, as opposed to an etiological or methodological goal. (2) The study outcome must be, or must contain the first event of ischemic stroke, and be assessed in a longitudinal design. (3) The study must present at least one measure to assess the predictive performance of the model. Studies in too distinct populations were excluded, such as studies on adverse outcomes of medical interventions and in-hospital stroke. Diagnostic algorithms and studies on genetic associations with ischemic stroke were excluded as well. The search strategy is explained in more detail in the **Supplementary material chapter 4, section 'search strategy'**.

4.2.1.2 Data extraction and risk of bias

Data extraction and quality assessment was conducted by Y.d.J. Included prediction models were assessed for risk of bias and applicability using the Prediction model Risk Of Bias ASsessment Tool (PROBAST)[15,16]. The PROBAST consists of 20 signaling questions for risk of bias within four domains (participant selection, predictors, outcome, and analysis) and three questions for applicability within the first three domains. In addition, we added the domain "usability", which describes whether the model could be used in the present form for risk prediction.

4.2.2 External validation

4.2.2.1 Study population and predictor definitions

The Netherlands Cooperative Study on the Adequacy of Dialysis (NECOSAD) was a prospective, multicenter cohort study in which 38 dialysis centers participated. Between 1997 and 2007, patients older than 18 years without previous renal replacement therapy were included at initiation of dialysis, which was defined as the baseline. Patients were censored when they underwent renal transplantation, died, or withdrew from the study. Although information on death and transplantation of NECOSAD is updated biannually (last update on 04-2019), information on stroke was available until 06-2009, which was used as censoring date. Weight and blood pressure were measured after dialysis. Medication usage and medical history were taken from patients charts. Smoking behavior was recorded as never, ceased, or current smoker. Cholesterol levels were measured in venous blood, and proteinuria was measured in 24 h urine sampling. For the external validation, we used the original predictor definition of the included studies if possible or selected a proxy based on literature and

clinical expertise. As the predictive performance is likely influenced by a less-stringent proxy selection, the model was excluded for validation if more than one predictor was different in NECOSAD compared with the original study.

4.2.2.2 Outcome

Our outcome ischemic stroke was defined as an ischemic cerebrovascular accident (CVA) requiring hospitalization, or fatal ischemic stroke. This was recorded in the study follow-up forms as CVA, which included both ischemic and hemorrhagic events. To exclude other diagnoses than ischemic stroke, such as hemorrhagic stroke, transient ischemic attack, or thromboembolisms, we developed key word searches in free text entries that were associated with hospitalizations, surgeries, and reasons for dialysis abatement. Furthermore, we used information from a subset of NECOSAD that was chart-reviewed as part of a data quality check.

4.2.2.3 Statistical analysis

For discrimination, the area under the ROC curve and Harrell's c-statistic for logistic and Cox regression models, respectively, were calculated. For calibration, we calculated the observed risk within the study's original timeframe using Kaplan-Meier survival probabilities for Cox models. Calibration plots were calculated using observed versus predicted probabilities in 10 equal-sized groups, and by fitting a LOWESS curve on the observed and predicted probabilities[17]. For models presenting event rates, the cumulative incidence was approximated (method detailed in the **Supplementary material chapter 4**). For models presenting only beta's without baseline risk, we estimated the constant by refitting the prognostic index (as these were all logistic models)[18]. Stroke prediction models were validated within the original timeframe if applicable, within the maximal follow-up if no timeframe was specified, or pragmatically within 10 years if both the timeframe and maximum follow-up were not specified. Missing data were assumed to be missing at random and were imputed using multiple imputation (detailed in **Supplementary material chapter 4**). We conducted four sensitivity analyses: 1. to further differentiate between ischemic and hemorrhagic stroke, we conducted a chart review as part of a data quality check. Of the 38 participating centers in NECOSAD, data from a representative subset of six dialysis centers (four regional hospitals and two academic hospitals, with a total of 755 patients; 38.6% of whole study sample) were chart-reviewed and model performance was subsequently evaluated in this cohort. 2. As vitamin K antagonists (VKA) may be prescribed for prevention of ischemic stroke in patients with a high risk of ischemic stroke, we performed an analysis only on those patients without VKA. 3. To estimate the effect of competing risk, a "worst-case" analysis

in which all patients that died were regarded as ischemic stroke was conducted as well. 4. Stratification on treatment modality, that is, hemodialysis and peritoneal dialysis. RStudio version 1.1.463 and IBM SPSS 25.0 were used.

4.3 RESULTS

4.3.1 Systematic search and study selection

The search yielded 901 references, of which 61 studies were included. Cross-reference searching resulted in an additional 16 studies. Of these 77 studies, 15 studies were subsequently validated (**Figure 4.1**). We validated 11 models with the exact same predictor definition as the original models; for the other four models, a proxy was used for one of the predictors in the model: gastrointestinal disease instead of history of bleeding for the GARFIELD-AF[19], a proxy that was used before[11]. For the model of Lip *et al.*[20], we used the whole follow-up time for the predictor "time within therapeutic range" if the patient used a VKA. Left ventricular hypertrophy diagnosed by ECG was a predictor in two models[21,22]; in NECOSAD, this was based on the medical history. All validated models are presented in more detail in the **Supplementary material chapter 4**.

4.3.2 Study characteristics

The characteristics of the included studies are shown in **Table 4.1**. Apart from two studies[8,30] with a retrospective or case-control design, all studies used a prospective study design, either as a randomized controlled trial or observational study. Five studies were conducted in the general population[21,26,28,29,32], eight studies in atrial fibrillation cohorts[7,8,19,20,23,24,30,31], one in a cardiovascular risk population[27], and one in patients with atherosclerosis[25]. A prediction timeframe was stated in ten studies and ranged between 1 and 10 years. In total, thirteen models used Cox and two used logistic regression. Models were presented as a point-based risk score in nine studies, a full formula or beta's with intercept in three studies, a calculator in one study, a decision rule in one study, and beta's without a constant in one study. There was substantial risk for overfitting in most models: eight studies did not perform internal validation or used split-sample validation. Events per variable, another indication of overfitting, ranged between 1.5 and 1,106.9 and was below 20 in eight of the fifteen studies. In most studies, model performance was good, with the original c-statistics (available for 12/15 studies) ranging between 0.61 for the CHA_2DS_2-VASc[7] and 0.82 for the $CHADS_2$[8]. A measure of calibration was given in nine studies, generally showing good calibration.

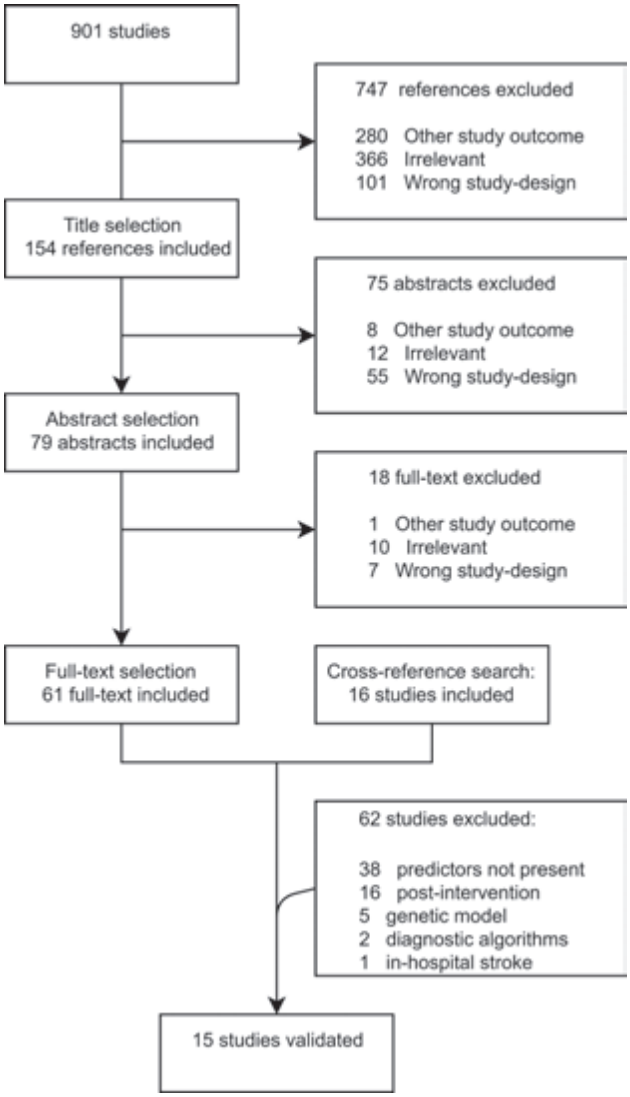


Figure 4.1 Flow chart of study selection. The label "irrelevant" was used for studies that did not present a prediction model; "other study outcome" was used for prediction studies that were not on ischemic stroke; "wrong study-design" was used for reviews on prediction studies, model updates, external validation, and implementation studies.

Study	Design	Outcome	Population	Male	Age	*n* total / *n* events	*n* cand. pred	Time frame	Model method	Internal validation	Discrimination	Calibration
				%		(%)	(EPV)				C-statistic	
AFI Investigators, 1994(30) Name: AFI	RCT	Ischemic stroke TIA Systemic embolus	AF	-	69*	3706 / 51 (1.38)	15 (3.4)	-	Cox	-	-	-
Gage, 2001(8) Name: $CHADS_2$	Cohort (retrospective)	Ischemic stroke TIA	AF	42.0	81*	1733 / 94 (5.42)	5 (14.2)	-	Cox	Bootstrapping	0.82	-
Wang, 2003(29) Name: Framingham heart study	Cohort (prospective)	Ischemic stroke Haemorrhagic stroke TIA	AF	53.2	75* (SD 9)	868 / 111 (12.79)	11 (10.1)	5y	Cox	Bootstrapping	0.66	Nam and D'Agostino
Chambless, 2004(32) Name: ARIC model	Cohort (prospective)	Ischemic stroke	Atherosclerosis	44.8	-	14 685 / 434 (2.96)	16 (27.1)	10y	Cox	Bootstrapping	-	Calibration curve
Zhang, 2005(27) Name: -	Cohort (prospective)	Ischemic stroke TIA	General	100	45* (SD 7.98)	3000 / 49 (1.63)	8 (6.1)	10y	Cox	Split-sample	0.72	Hosmer-Lemeshow Observed vs expected
Diener, 2005(31) Name: Essen stroke risk score	RCT	Stroke (undefined)	Cardiovascular	-	-	19 099 / 775 (3.95)	32 (24.2)	-	Cox	-	-	-
Wu, 2006(25) Name: -	Cohort (prospective)	Ischemic stroke CHD	General	49.4	46* (SD 6)	9903 / 371 (3.75)	6 (61.8)	10y	Cox	-	0.80	Hosmer-Lemeshow Observed vs expected
Assmann, 2007(24) Name: PROCAM Risk score	Cohort (prospective)	Stroke (undefined) TIA	General	72.6	45.7* (SD 6.8)	8130 / 85 (1.05)	57 (1.5)	10y	Cox	Cross-validation	0.78	Observed vs expected

Study	Design	Outcome	Population	Male	Age	*n* total / *n* events	*n* cand. pred	Time frame	Model method	Internal validation	Discrimination	Calibration
				%		(%)	(EPV)				C-statistic	
Rietbrock, 2008(23) Name: Modified-$CHADS_2$	Case-control	Ischemic stroke Haemorrhagic stroke	AF	48.6	-	305 566 / 19 925 (6.52)	18 (1106.9)	5y	Cox	-	0.72	-
Lip, 2010(7) Name: CHA_2DS_2-VASC	Cohort (prospective)	Ischemic stroke TIA Systemic embolus	AF	59.2	66* (SD 14)	1084 / 25 (2.31)	9 (2.8)	1y	Logistic	-	0.61	-
Lip, 2013(20) Name: -	RCT	Stroke (undefined) Major bleeding Systemic embolus	AF	65.0	70* (SD 9)	2293 / 94 (4.1)	6 (15.7)	-	Logistic	-	0.73	-
Singer, 2013(28) Name: ATRIA	RCT	Ischemic stroke Systemic embolus	AF	-	-	10 927 / 685 (6.27)	12 (57.1)	-	Cox	Split-sample Bootstrapping	0.73	Nam and D'Agostino
Yatsuya, 2013(26) Name: -	Cohort (prospective)	Ischemic stroke Hemorrhagic stroke	General	33.9	-	15 672 / 790 (5.04)	10 (79)	10y	Cox	Bootstrapping	0.73	Gronnesby and Borgan Observed vs expected
Ferket, 2014(33) Name: -	Cohort (prospective)	Ischemic stroke	General	42.6	-	27 493 / 2559 (9.31)	14 (182.8)	10y	Cox	Cross-validation Bootstrapping	0.76	Calibration curve Nam and D'Agostino
Fox, 2017(19) Name: GARFIELD-AF	Cohort (prospective)	Ischemic stroke Systemic embolus TIA	AF	55.5	71† (63-78)	38 935 / 473 (1.21)	32 (11.9)	1y	Cox	Cross-validation	0.69	Calibration curve Nam and D'Agostino

Table 4.1 Overview of the 15 included and externally validated studies. Abbreviations: RCT: randomized controlled trial; TIA: transient ischemic attack; AF: atrial fibrillation; CHD: coronary heart disease; SD: standard deviation; EPV: events per variable. For age, the values are: *mean, †median or not stated.

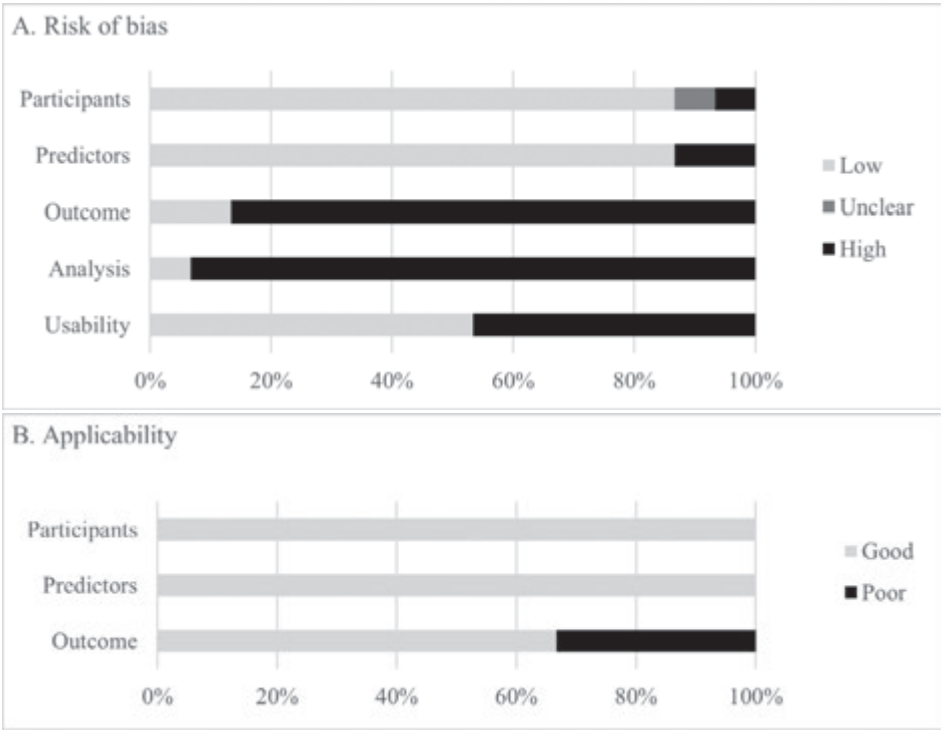


Figure 4.2 PROBAST Risk of bias summary, percentage of studies with a low-, unclear- and high risk of bias (figure 2A) and the percentage of studies with a good-, unclear- and poor applicability (figure 2B) per domain of the PROBAST tool. The domain 'usability', which was added by the authors of the present study, consisted of one question: "could the model be used in the present form for risk prediction?"

4.3.3 Risk of bias

All included studies showed high risk of bias on at least one domain of the PROBAST tool, with three studies scoring a high risk of bias on three domains (**Figure 4.2A**; details on individual studies are presented in **Supplementary material chapter 4, Table S1**). Thirteen studies scored poor on the domain "outcome", mainly because of the absence of a time interval, or the use of composite outcomes, such as "stroke" (which could include combinations of ischemic stroke, hemorrhagic stroke, or TIA). Of the 15 studies, fourteen scored poor on the domain "analysis", which included predictor selection, competing risks, overfitting, and model performance. Only one model accounted for competing risks for ischemic stroke, namely intracranial hemorrhage and death from other causes[33]. The applicability of the models was generally good (**Figure 2B; Supplementary material chapter 4, Table S1**). Seven models were less or not applicable for individual risk prediction: for six models, this was because no predicted probability was given, but an observed event rate[7,8,23,27,30,31], for one model[20] beta's were given but no constant was provided.

4.3.4 Validation cohort

The baseline characteristics of NECOSAD are presented in **Table 4.2**. In total, 2,051 patients were enrolled, of which 1,955 (95.3%) were followed after the baseline measurements and subsequently used for the present study. The mean age was 59.98 year (SD 15.1) and 1,216 (62.2%) patients were male. Most patients were on hemodialysis (64.9%), the remainder on peritoneal dialysis. At the end of a median follow-up of 2.5 year, 127 (6.5%) patients suffered an ischemic stroke, 43 of which were fatal. A total of 846 (43.3%) patients died on other causes during follow-up, whereas 571 (29.2%) patients received a transplant.

	Total N=1955 (%)	**Missing N (%)**	**No stroke N= 1828 (%)**	**Stroke N=127 (%)**
Age (mean, SD)	59.98 (15.1)	3 (0.2)	59.59 (15.2)	65.56 (12.1)
Sex (Male, %)	1216 (62.2)	4 (0.2)	1147 (62.7)	69 (54.3)
Vitamin K antagonist use, %	221 (11.3)	210 (10.7)	200 (10.9)	21 (16.8)
Antiplatelet drug use, %	396 (20.3)	210 (10.7)	355 (19.4)	41 (32.3)
Antihypertensive drug use, %	1439 (73.6)	210 (10.7)	1335 (73.0)	104 (81.9)
Systolic blood pressure >140 mmHg, %	1090 (55.8)	20	1024 (56.0)	66 (52.0)
Smoking		200 (10.2)		
Current	392 (20.1)		367 (20.1)	25 (19.7)
Ever	792 (40.5)		734 (40.2)	58 (45.7)
Comorbidities,%				
Prior stroke	146 (7.5)	193 (9.9)	126 (6.9)	20 (15.7)
Heart failure	201 (10.3)	193 (9.9)	182 (10.0)	19 (15.0)
Left ventricle hypertrophy	258 (13.2)	193 (9.9)	229 (12.5)	29 (22.8)
Peripheral artery disease	245 (12.5)	193 (9.9)	215 (11.8)	30 (23.6)
Coronary artery disease	193 (9.9)	193 (9.9)	176 (9.6)	17 (13.4)
Malignancy	169 (8.6)	194 (9.9)	161 (8.8)	8 (6.3)
Diabetes	387 (19.8)	193 (9.9)	354 (19.4)	33 (26.0)
Dialysis modality (%)		10 (0.5)		
Haemodialysis	1268 (64.9)		1177 (64.4)	91 (71.7)
Peritoneal dialysis	677 (34.6)		642 (35.1)	35 (27.6)
Primary kidney disease, (%)		192 (9.8)		
Diabetic nephropathy	281 (14.4)		254 (13.9)	27 (21.3)
Glomerulonephritis	240 (12.3)		229 (12.5)	11 (8.7)
Vascular	331 (16.9)		291 (15.9)	40 (31.5)
Other	911 (46.6)		863 (47.2)	48 (37.8)

Table 4.2 Characteristics of validation cohort NECOSAD before multiple imputation.

4.3.5 Performance of stroke risk scores

4.3.5.1 Discrimination

While the discrimination of the original studies was moderate to good, it was poor in the validation cohort, with c-statistics ranging from 0.49 (95% CI 0.40–0.58) for the study by Wu[28] to 0.66 (0.59–0.74) for the GARFIELD-AF model[19]. Except for the CHA_2DS_2-VASc[7], all models that presented a c-statistic in the original study were less able to discriminate between low- and high-risk patients in the validation cohort (**Table 4.3**). These results were consistent in the sensitivity analyses: recoding all patients who died as ischemic stroke instead of censoring increased discrimination of all models slightly with c-statistics increasing on average with 0.04 (range 0 to 0.08). Stratifying between treatment modality increased discrimination marginally for hemodialysis (c-statistic average increase 0.02, range −0.01 to 0.07), and decreased for peritoneal dialysis (average −0.06, range −0.13 to −0.03). For the other sensitivity analyses, discrimination was consistent with the main analysis: in the chart-reviewed patients, the average difference with the main analysis was 0 (range −0.07 to 0.14); in non-VKA users, this was also 0 (−0.01 to 0.02), detailed in **Supplementary material chapter 4, Table S6 and Figure S2**.

4.3.5.2 Calibration

Calibration plots are shown in **Figure 4.3**. Predicted probabilities for six models[7,8,23,27,30,31] were approximated, as only event rates were given in the original studies. The Framingham Heart Study[24] was the only study showing good calibration. For the other studies, calibration was poor both in respect to the actual agreement between observed and predicted probabilities (calibration-in-the-large, **Table 4.3**) as well as the calibration curves, which showed over- or underprediction, or a combination of both. The broadness of the range of predicted probabilities differed between studies: 0.05% to 6.61% for Zhang *et al.*[26] and 0.046% to 92.13% for Chambless *et al.*[25]. Calibration was comparable in the sensitivity analyses, but as the observed risk was notably higher in sensitivity analysis three in which death was recoded to ischemic stroke, calibration differed more substantially (**Supplementary material chapter 4, Table S7, Figures S3–S8**). Models with a short timeframe did not perform differently, nor did models for which a proxy predictor was used[19,20,21,22].

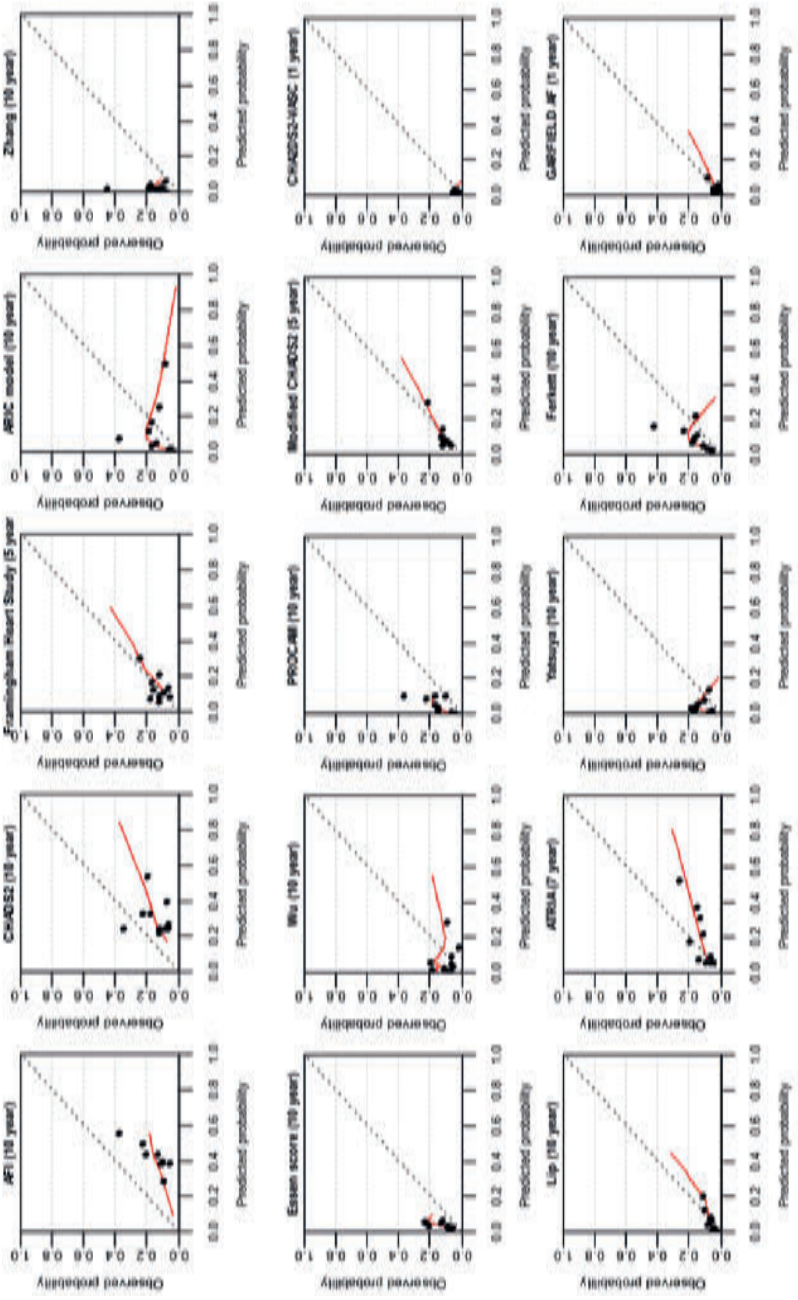


Figure 4.3 Calibration plots of the included studies, showing observed and predicted probabilities for ischemic stroke in NECOSAD. 6 studies (AFI, $CHADS_2$, Essen score, modified $CHADS_2$ and the CHA_2DS_2-VASC) provided event rates, which were recalculated to cumulative incidences. The model by Lip was presented without a constant, and was subsequently recalibrated.

Study	Model name	Model type	Discrimination		Calibration	
			Original C-statistic	Validation C-statistic (95% CI)	Observed	Predicted
AFI Investigators, 1994(30)	AFI	Decision rules	-	0.61 (0.56-0.65)	0.1638	0.4154
Gage, 2001(8)	$CHADS_2$	Risk score	0.82	0.61 (0.56-0.66)	0.1638	0.3058
Wang, 2003(29)	Framingham Heart Study	Risk score	0.66	0.57 (0.50-0.63)	0.1379	0.1260
Chambless, 2004(32)	ARIC	Formula	-	0.61 (0.56-0.66)	0.1638	0.1247
Zhang, 2005(27)	-	Formula	0.72	0.53 (0.48-0.58)	0.1638	0.0243
Diener, 2005(31)	Essen stroke risk score	Risk score	-	0.64 (0.59-0.70)	0.1638	0.0428
Wu, 2006(25)	-	Risk score	0.80	0.49 (0.40-0.58)	0.0926	0.0759
Assmann, 2007(24)	PROCAM	Risk score	0.78	0.61 (0.56-0.66)	0.1638	0.0515
Rietbrock, 2008(23)	Modified-CHADS2	Risk score	0.72	0.62 (0.56-0.68)	0.1153	0.1032
Lip, 2010(7)	CHA_2DS_2-VASC	Risk score	0.61	0.65 (0.57-0.73)	0.0256	0.0187
Lip, 2013(20)	-	Betas (no constant)	0.73	0.60 (0.54-0.65)	NA*	NA*
Singer, 2013(28)	ATRIA	Risk score	0.73	0.63 (0.58-0.69)	0.1215	0.1965
Yatsuya, 2013(26)	-	Risk score	0.73	0.56 (0.50-0.63)	0.1113	0.0472
Ferket, 2014(33)	-	Calculator	0.76	0.61 (0.56-0.66)	0.1638	0.0933
Fox, 2017(19)	GARFIELD-AF	Formula	0.69	0.66 (0.59-0.74)	0.0277	0.0372

Table 4.3 Predictive performance of the 15 included studies, in the original study and in the NECOSAD external validation cohort. *The model by Lip et al. was provided without a constant and was recalibrated for the external validation, resulting in values for observed and predicted probabilities that are equal.

4.4 DISCUSSION

In this prospective cohort study of 1,955 incident dialysis patients, we externally and independently validated 15 predictive models for ischemic stroke. All studies showed poor predictive performance, both for discrimination and calibration. C-statistics ranged between 0.49 for the model by Wu *et al*[28] and 0.66 for the GARFIELD-AF[19], where a c-statistic >0.80 is usually regarded as good. Apart from the Framingham Heart Study[24] which was well calibrated, calibration was also poor. External validation of

the $CHADS_2$ and CHA_2DS_2-VASc has only twice been performed in prevalent dialysis patients, yielding comparable results with our study in a large study on 10.999 atrial fibrillation patients on dialysis[9]. However, this study was in prevalent dialysis patients only. The second study showed better predictive performance, but was conducted in a small sample of 141 atrial fibrillation patients on dialysis with only 15 events[10]. Both studies presented discrimination but offered no information on calibration, which, with regard to risk comparison between bleeding and ischemic stroke, could be argued to be of more importance than discrimination.

The poor predictive performance in dialysis patients could have several explanations. First, we demonstrated the high risk of bias. For example, while the rule of thumb of 10 or 20 events per variable is debated[34,35,36,37], and more nuanced methods exists[38,39], it is generally accepted that a lower number of events per predictor may result in overfitting and consequently reduce the external validity[15,40]. In our study, more than half of the included studies used less than 20 events per predictor. This observation is not unique to models on ischemic stroke[41], but is demonstrated in other fields as well, for example, for models predicting end-stage renal disease in patients with chronic kidney disease[42]. The TRIPOD guidelines[13] and the PROBAST tool[15,16], both recently published, can possibly aid authors developing new prediction models to avoid commonly encountered methodological errors. Second, differences in case-mix heterogeneity between the original development cohort and the external validation cohort may result in lower discriminative ability even if the fitted regression coefficients are correct[43]. As NECOSAD is likely a more homogeneous cohort than the development cohorts of the validated models, reduced discriminative ability may partly be explained by case-mix difference. Furthermore, predictors that have predictive value in the original development cohort may be of less value in dialysis cohorts due to patient characteristics. Other, more dialysis-specific predictors may be better able to discriminate in this relatively homogeneous population. However, it should be noted that these models are commonly used in incident dialysis patients and thus reflect the actual predictive performance in current clinical use. Third, competing risks (e.g. transplantation, cessation of dialysis therapy, death, or loss to follow-up) may play a major role and greatly impact predictive performance[44]. In our 15 validated studies, only one study accounted for such competing risks[33]. To demonstrate the possible effect of competing risk, we performed a "worst case" sensitivity analysis in which all patients who died were regarded as having the outcome as well. While the discrimination of all models showed a modest increase, calibration was off, as all models underpredicted this artificially increased risk.

The main strength of this study is the independent external validation of 15 different ischemic stroke risk models in the same population of incident dialysis patients, allowing comparison

of models and increasing the number of validated models in this clinically relevant population substantially. Furthermore, the large and well-defined prospective cohort of 1,955 incident dialysis patients, with a substantial number of events allowed for well-powered analyses at a clinical relevant time point, namely the initiation of dialysis. Our study has several limitations. First, while CVA was recorded in NECOSAD, this included both hemorrhagic and ischemic events. We developed two strategies to overcome this problem: first, we searched for text entries that differentiated between hemorrhagic and ischemic CVA. Second, more than a third of the patients were chart-reviewed and used as data-quality check. Validating the models in this subset resulted in similar predictive performance, but with a higher degree of uncertainty due to the reduced sample size and lower number of events. Another limitation is the lack of information on cardiac arrhythmias, such as atrial fibrillation, limiting the number of possible models to validate. Using prescription of oral anticoagulation as proxy for atrial fibrillation seems reasonable, and was done in our previous study[11]. However, as oral anticoagulation is directly and protectively associated with ischemic stroke, we refrained from validating these models in our cohort. As atrial fibrillation is the main indication for VKA use in dialysis patients, we performed a sensitivity analysis in the non-VKA users, the results of which were similar to the main analysis. We considered to perform the same analysis in VKA users only, but refrained from doing so because of the low number of patients and events in this subgroup. Nevertheless, analyzing the performance in all dialysis patients rather than in a subgroup of patients with atrial fibrillation is clinically relevant because many dialysis patients have an atrial fibrillation event without being diagnosed[45]. Finally, we validated all models for the single-outcome ischemic stroke. As most models predicted a composite outcome, this may have reduced the predictive performance.

Kidney disease and stroke share common risk factors, such as hypertension, aging, diabetes mellitus, and dyslipidemia[46]. The occurrence of atrial fibrillation, also a major risk factor for ischemic stroke in dialysis patients, is more than ten times frequent compared with the general population[47,48,49,50], is increasing in prevalence[51] and is often unnoticed[45]. Other risk factors include volumetric changes associated with both end-stage renal disease and dialysis therapy[52,53,54], and the accelerated atherosclerotic cerebral vascular disease caused in part by the uremic process[52,53,54, 55,56,57], and the accelerated atherosclerotic cerebral vascular disease caused in part by the uremic process[55,56,57]. Apart from the increased risk of stroke in dialysis patients, the risk of hemorrhage is also increased. Until now, no randomized controlled trials on stroke prevention with any form of anticoagulation have been performed in dialysis patients[58]. No high-quality guidelines on stroke prevention in this population exists. Furthermore, we have previously shown that commonly used bleeding risk models have poor predicting performance in incident dialysis patients and

should not be used in this population[11]. The poor predictive performance of 15 ischemic stroke risk models of this present study is complementary to these findings. Thus, while the use of these clinical decision aids are appealing as a method to standardize the allocation of care in a seemingly objective manner, clinicians should keep these limitations in mind when applying these models and also consider more dialysis specific variables.

In summary, we have demonstrated the poor predictive performance of ischemic stroke risk models in dialysis patients in addition to our recent external validation of bleeding risk scores. These notions warrant caution for risk stratification in dialysis patients and underline the urgent need for prediction model development specifically targeted at dialysis patients. Alternatively, promising existing models, such as the Framingham Heart Score, which showed good calibration but poor discrimination, could be updated by incorporating dialysis-specific variables.

4.5 ACKNOWLEDGMENTS

The Netherlands Cooperative Study on the Adequacy of Dialysis (NECOSAD) was funded in part by an unrestricted grant from the Dutch Kidney Foundation. The funding source was involved in neither the collection, interpretation, and analysis of the data nor the decision on the writing and submission of this report for publication. This present study was supported by a grant from the Dutch Kidney Foundation (16OKG12).

4.6 REFERENCES

1. V.L. Feigin, B. Norrving, G.A. Mensah Global burden of stroke Circ Res, 120 (2017), pp. 439-448
2. M. Kawamura, S. Fijimoto, S. Hisanaga, Y. Yamamoto, T. Eto Incidence, outcome, and risk factors of cerebrovascular events in patients undergoing maintenance hemodialysis Am J Kidney Dis, 31 (1998), pp. 991-996
3. S.L. Seliger, D.L. Gillen, W.T. Longstreth Jr., B. Kestenbaum, C.O. Stehman-Breen Elevated risk of stroke among patients with end-stage renal disease Kidney Int, 64 (2003), pp. 603-609
4. K. Toyoda, K. Fujii, T. Ando, Y. Kumai, S. Ibayashi, M. Iida Incidence, etiology, and outcome of stroke in patients on continuous ambulatory peritoneal dialysis Cerebrovasc Dis, 17 (2004), pp. 98-105
5. Power, K. Chan, S.K. Singh, D. Taube, N. Duncan Appraising stroke risk in maintenance hemodialysis patients: a large single-center cohort study Am J Kidney Dis, 59 (2012), pp. 249-257
6. Ovbiagele Chronic kidney disease and risk of death during hospitalization for stroke J Neurol Sci, 301 (2011), pp. 46-50
7. G.Y. Lip, R. Nieuwlaat, R. Pisters, D.A. Lane, H.J. Crijns Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation Chest, 137 (2010), pp. 263-272
8. B.F. Gage, A.D. Waterman, W. Shannon, M. Boechler, M.W. Rich, M.J. Radford Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation JAMA, 285 (2001), pp. 2864-2870
9. T.F. Chao, C.J. Liu, K.L. Wang, Y.J. Lin, S.L. Chang, L.W. Lo, *et al.* Incidence and prediction of ischemic stroke among atrial fibrillation patients with end-stage renal disease requiring dialysis Heart Rhythm, 11 (2014), pp. 1752-1759
10. T.K. Wang, J. Sathananthan, M. Marshall, A. Kerr, C. Hood Relationships between anticoagulation, risk scores and adverse outcomes in dialysis patients with atrial fibrillation Heart Lung Circ, 25 (2016), pp. 243-249
11. G. Ocak, C. Ramspek, M.B. Rookmaaker, P.J. Blankestijn, M.C. Verhaar, W.J.W. Bos, *et al.* Performance of bleeding risk scores in dialysis patients Nephrol Dial Transplant, 34 (2019), pp. 1223-1231
12. A. Liberati, D.G. Altman, J. Tetzlaff, C. Mulrow, P.C. Gotzsche, J.P. Ioannidis, *et al.*The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration BMJ, 339 (2009), p. b2700
13. G.S. Collins, J.B. Reitsma, D.G. Altman, K.G. Moons Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement BMJ, 350 (2015), p. g7594
14. K.G. Moons, J.A. de Groot, W. Bouwmeester, Y. Vergouwe, S. Mallett, D.G. Altman, *et al.* Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist PLoS Med, 11 (2014), p. e1001744
15. K.G.M. Moons, R.F. Wolff, R.D. Riley, P.F. Whiting, M. Westwood, G.S. Collins, *et al.* PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration Ann Intern Med, 170 (2019), pp. W1-W33
16. R.F. Wolff, K.G.M. Moons, R.D. Riley, P.F. Whiting, M. Westwood, G.S. Collins, *et al.* A tool to assess the risk of bias and applicability of prediction model studies Ann Intern Med, 170 (2019), pp. 51-58

17. W.S. Cleveland LOWESS: a program for smoothing scatterplots by robust locally weighted regression Am Stat, 35 (1981), p. 54

18. K.J. Janssen, Y. Vergouwe, C.J. Kalkman, D.E. Grobbee, K.G. Moons A simple method to adjust clinical prediction models to local circumstances Can J Anaesth, 56 (2009), pp. 194-201

19. K.A.A. Fox, J.E. Lucas, K.S. Pieper, J.P. Bassand, A.J. Camm, D.A. Fitzmaurice, *et al.* Improved risk stratification of patients with atrial fibrillation: an integrated GARFIELD-AF tool for the prediction of mortality, stroke and bleed in patients with and without anticoagulation BMJ Open, 7 (2017), p. e017157

20. G.Y.H. Lip, D.A. Lane, H. Buller, S. Apostolakis Development of a novel composite stroke and bleeding risk score in patients with atrial fibrillation: the AMADEUS Study Chest, 144 (2013), pp. 1839-1847

21. R.B. D'Agostino, P.A. Wolf, A.J. Belanger, W.B. Kannel Stroke risk profile: adjustment for antihypertensive medication. The Framingham Study Stroke, 25 (1994), pp. 40-43

22. P.A. Wolf, R.B. D'Agostino, A.J. Belanger, W.B. Kannel Probability of stroke: a risk profile from the Framingham Study Stroke, 22 (1991), pp. 312-318

23. Atrial Fibrillation Investigators Risk factors for stroke and efficacy of antithrombotic therapy in atrial fibrillation. Analysis of pooled data from five randomized controlled trials Arch Intern Med, 154 (1994), pp. 1449-1457

24. T.J. Wang, J.M. Massaro, D. Levy, R.S. Vasan, P.A. Wolf, R.B. D'Agostino, *et al.* A risk score for predicting stroke or death in individuals with new-onset atrial fibrillation in the community: the Framingham Heart Study JAMA, 290 (2003), pp. 1049-1056

25. L.E. Chambless, G. Heiss, E. Shahar, M.J. Earp, J. Toole Prediction of ischemic stroke risk in the atherosclerosis risk in communities study Am J Epidemiol, 160 (2004), pp. 259-269

26. X.F. Zhang, J. Attia, C. D'Este, X.H. Yu, X.G. Wu A risk score predicted coronary heart disease and stroke in a Chinese cohort J Clin Epidemiol, 58 (2005), pp. 951-958

27. H.C. Diener, P.A. Ringleb, P. Savi Clopidogrel for the secondary prevention of stroke Expert Opin Pharmacother, 6 (2005), pp. 755-764

28. Y. Wu, X. Liu, X. Li, Y. Li, L. Zhao, Z. Chen, *et al.* Estimation of 10-year risk of fatal and nonfatal ischemic cardiovascular diseases in Chinese adults Circulation, 114 (2006), pp. 2217-2225

29. G. Assmann, H. Schulte, P. Cullen, U. Seedorf Assessing risk of myocardial infarction and stroke: new data from the prospective cardiovascular munster (PROCAM) study Eur J Clin Invest, 37 (2007), pp. 925-932

30. S. Rietbrock, E. Heeley, J. Plumb, T. van Staa Chronic atrial fibrillation: incidence, prevalence, and prediction of stroke using the Congestive heart failure, Hypertension, Age >75, Diabetes mellitus, and prior Stroke or transient ischemic attack (CHADS2) risk stratification scheme Am Heart J, 156 (2008), pp. 57-64

31. D.E. Singer, Y. Chang, L.H. Borowsky, M.C. Fang, N.K. Pomernacki, N. Udaltsova, *et al.* A new risk scheme to predict ischemic stroke and other thromboembolism in atrial fibrillation: the ATRIA study stroke risk score J Am Heart Assoc, 2 (2013), p. e000250

32. H. Yatsuya, H. Iso, K. Yamagishi, Y. Kokubo, I. Saito, K. Suzuki, *et al.* Development of a point-based prediction model for the incidence of total stroke: Japan public health center study Stroke, 44 (2013), pp. 1295-1302

33. B.S. Ferket, B.J. van Kempen, R.G. Wieberdink, E.W. Steyerberg, P.J. Koudstaal, A. Hofman, *et al.* Separate prediction of intracerebral hemorrhage and ischemic stroke Neurology, 82 (2014), pp. 1804-1812

34. P. Peduzzi, J. Concato, A.R. Feinstein, T.R. Holford Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates J Clin Epidemiol, 48 (1995), pp. 1503-1510

35. E. Vittinghoff, C.E. McCulloch Relaxing the rule of ten events per variable in logistic and Cox regression Am J Epidemiol, 165 (2007), pp. 710-718

36. M. van Smeden, J.A. de Groot, K.G. Moons, G.S. Collins, D.G. Altman, M.J. Eijkemans, *et al.* No rationale for 1 variable per 10 events criterion for binary logistic regression analysis BMC Med Res Methodol, 16 (2016), p. 163

37. D.S. Courvoisier, C. Combescure, T. Agoritsas, A. Gayet-Ageron, T.V. Perneger Performance of logistic regression modeling: beyond the number of events per variable, the role of data structure J Clin Epidemiol, 64 (2011), pp. 993-1000

38. M. van Smeden, K.G. Moons, J.A. de Groot, G.S. Collins, D.G. Altman, M.J. Eijkemans, *et al.* Sample size for binary logistic prediction models: beyond events per variable criteria Stat Methods Med Res, 28 (2019), pp. 2455-2474

39. R.D. Riley, K.I. Snell, J. Ensor, D.L. Burke, F.E. Harrell Jr., K.G. Moons Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes Stat Med, 38 (2019), pp. 1276-1296

40. P.C. Austin, E.W. Steyerberg Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models Stat Methods Med Res, 26 (2017), pp. 796-808

41. W. Bouwmeester, N.P. Zuithoff, S. Mallett, M.I. Geerlings, Y. Vergouwe, E.W. Steyerberg, *et al.* Reporting and methods in clinical prediction research: a systematic review PLoS Med, 9 (2012), pp. 1-12

42. C.L. Ramspek, Y. de Jong, F.W. Dekker, M. van Diepen Towards the best kidney failure prediction tool: a systematic review and selection aid Nephrol Dial Transplant (gfz018) (2019), 10.1093/ndt/gfz018

43. Y. Vergouwe, K.G. Moons, E.W. Steyerberg External validity of risk models: use of benchmark values to disentangle a case-mix effect from incorrect coefficients. Am J Epidemiol, 172 (2010), pp. 971-980

44. C.J. Shih, S.M. Ou, P.W. Chao, S.C. Kuo, Y.J. Lee, C.Y. Yang, *et al.* Risks of death and stroke in patients undergoing hemodialysis with new-onset atrial fibrillation: a competing-risk analysis of a nationwide cohort Circulation, 133 (2016), pp. 265-272

45. M.S. Buiten, M.K. de Bie, J.I. Rotmans, B.A. Gabreels, W. van Dorp, R. Wolterbeek, *et al.* The dialysis procedure as a trigger for atrial fibrillation: new insights in the development of atrial fibrillation in dialysis patients Heart, 100 (2014), pp. 685-690

46. K. Toyoda, T. Ninomiya Stroke and cerebrovascular diseases in patients with chronic kidney disease Lancet Neurol, 13 (2014), pp. 823-833

47. O. Konigsbrugge, F. Posch, M. Antlanger, J. Kovarik, R. Klauser-Braun, J. Kletzmayr, *et al.* Prevalence of atrial fibrillation and antithrombotic therapy in hemodialysis patients: cross-sectional results of the Vienna InVestigation of AtriaL fibrillation and thromboembolism in patients on HemoDIalysis (VIVALDI) PLoS One, 12 (2017), p. e0169400

48. S. Genovesi, D. Pogliani, A. Faini, M.G. Valsecchi, A. Riva, F. Stefani, *et al.* Prevalence of atrial fibrillation and associated factors in a population of long-term hemodialysis patients Am J Kidney Dis, 46 (2005), pp. 897-902

49. E. Vazquez, C. Sanchez-Perales, F. Garcia-Garcia, P. Castellano, M.J. Garcia-Cortes, A. Liebana, *et al.* Atrial fibrillation in incident dialysis patients Kidney Int, 76 (2009), pp. 324-330

50. W.C. Winkelmayer, A.R. Patrick, J. Liu, M.A. Brookhart, S. Setoguchi The increasing prevalence of atrial fibrillation among hemodialysis patients J Am Soc Nephrol, 22 (2011), pp. 349-357

51. A.O. Molnar, M.M. Sood Predicting in a predicament: stroke and hemorrhage risk prediction in dialysis patients with atrial fibrillation Semin Dial, 31 (2018), pp. 37-47

52. D. Gottlieb, B. Mildworf, D. Rubinger, E. Melamed The regional cerebral blood Flow in patients under chronic hemodyalytic treatmentJ Cereb Blood Flow Metab, 7 (1987), pp. 659-661

53. A. Postiglione, F. Faccenda, G. Gallotta, P. Rubba, S. Federico Changes in middle cerebral artery blood velocity in uremic patients after hemodialysis Stroke, 22 (1991), pp. 1508-1511

54. R. Hata, M. Matsumoto, N. Handa, H. Terakawa, Y. Sugitani, T. Kamada Effects of hemodialysis on cerebral circulation evaluated by transcranial Doppler ultrasonography Stroke, 25 (1994), pp. 408-412

55. R. Kennedy, C. Case, R. Fathi, D. Johnson, N. Isbel, T.H. Marwick Does renal failure cause an atherosclerotic milieu in patients with end-stage renal disease? Am J Med, 110 (2001), pp. 198-204

56. K. Amann, R. Neususs, E. Ritz, T. Irzyniec, G. Wiest, G. Mall Changes of vascular architecture independent of blood pressure in experimental uremia Am J Hypertens, 8 (1995), pp. 409-417

57. T. Kawagishi, Y. Nishizawa, T. Konishi, K. Kawasaki, M. Emoto, T. Shoji, *et al.* High-resolution B-mode ultrasonography in evaluation of atherosclerosis in uremia Kidney Int, 48 (1995), pp. 820-826

58. T.F. Chao, S.A. Chen Risk of ischemic stroke and stroke prevention in patients with atrial fibrillation and renal dysfunction J Atr Fibrillation, 8 (2015), p. 1196

Chapter 5

VALIDATION OF RISK SCORES FOR ISCHEMIC STROKE IN ATRIAL FIBRILLATION ACROSS THE SPECTRUM OF KIDNEY FUNCTION

Ype de Jong
Edouard L. Fu
Merel van Diepen
Marco Trevisan
Karolina Szummer
Friedo W. Dekker
Juan J. Carrero
Gurbey Ocak

European Heart Journal 2021 Apr 14;42(15):1476-1485.
DOI: 10.1093/eurheartj/ehab059.
PMID: 33769473;
PMCID: PMC8046502.

To minimize the impact of printing on the environment, the supplementary material has been uploaded. For Chapter 5, the supplement is 33 pages, and can be downloaded using the QR code.

Abstract

Aims: The increasing prevalence of ischemic stroke (IS) can partly be explained by the likewise growing number of patients with chronic kidney disease (CKD). Risk scores have been developed to identify high-risk patients, allowing for personalized anticoagulation therapy. However, predictive performance in CKD is unclear. The aim of this study is to validate six commonly used risk scores for IS in atrial fibrillation (AF) patients across the spectrum of kidney function.

Methods and results: Overall, 36 004 subjects with newly diagnosed AF from SCREAM (Stockholm CREAtinine Measurements), a healthcare utilization cohort of Stockholm residents, were included. Predictive performance of the AFI, $CHADS_2$, Modified $CHADS_2$, CHA_2DS_2-VASc, ATRIA, and GARFIELD-AF risk scores was evaluated across three strata of kidney function: normal kidney function [estimated glomerular filtration rate (eGFR) >60 mL/min/1.73 m^2], mild CKD (eGFR 30–60 mL/min/1.73 m^2), and advanced CKD (eGFR <30 mL/min/1.73 m^2). Predictive performance was assessed by discrimination and calibration. During 1.9 years, 3069 (8.5%) patients suffered an IS. Discrimination was dependent on eGFR: the median c-statistic in normal eGFR was 0.75 (range 0.68–0.78), but decreased to 0.68 (0.58–0.73) and 0.68 (0.55–0.74) for mild and advanced CKD, respectively. Calibration was reasonable and largely independent of eGFR. The Modified $CHADS_2$ score showed good performance across kidney function strata, both for discrimination [c-statistic: 0.78 (95% confidence interval 0.77–0.79), 0.73 (0.71–0.74) and 0.74 (0.69–0.79), respectively] and calibration.

Conclusion: In the most clinically relevant stages of CKD, predictive performance of the majority of risk scores was poor, increasing the risk of misclassification and thus of over- or undertreatment. The Modified $CHADS_2$ score performed good and consistently across all kidney function strata, and should therefore be preferred for risk estimation in AF patients.

5.1 INTRODUCTION

The prevalence of ischemic stroke (IS) is increasing and has become a leading cause of morbidity and mortality worldwide.[1] Chronic kidney disease (CKD) is associated with an increased risk of IS via various mechanisms, both specific to CKD (e.g. accelerated atherosclerotic vascular disease) and general risk factors, such as hypertension, diabetes mellitus, dyslipidemia, and ageing.[2] With an estimated prevalence of 10–15% in the general population, a number that is increasing steadily,[3] CKD may partly explain the high number of strokes.[4,5] Atrial fibrillation (AF), which is considered the main risk factor for IS both in the general population and in CKD patients, is more commonly reported in this fragile population, an observation that may be related to shared risk factors such as age, diabetes, and hypertension.[3,6–8]

Risk scores for IS are essential to weigh the risk of IS vs. the risk of treatment-related bleeding and thus deliver patient-tailored therapy. In patients with CKD, this notion is highly relevant since these patients are at increased risk of treatment-related bleeding as well.[9–11] Typically, most risk scores use clinical parameters (e.g. disease history) in combination with patient-specific characteristics (e.g. age and sex) to compute a risk for IS within a given prediction timeframe. Although widely used risk scores, such as $CHADS_2$ and its updated version CHA_2DS_2-VASc, are endorsed by current guidelines on IS,[8,12–14] their predictive performance in patients with CKD is largely unknown, as these risk scores have been developed in general AF populations.[2,8] For incident dialysis patients, however, external validation studies showed poor predictive performance both for risk scores predicting IS and bleeding.[15,16] Despite these uncertainties, the use of these clinical decision aids is appealing as a seemingly objective tool to standardize the allocation of anticoagulation therapy within CKD care. However, due to the lack of information on the validity of these risk scores in patients beyond the development cohorts of the original studies, their use comes with a risk of misclassification. The aim of the present study is therefore to externally validate multiple commonly used risk scores for IS in a cohort of patients with AF across the spectrum of kidney function.

5.2 METHODS

This study was reported in line with the TRIPOD guideline.[17]

5.2.1 Study population and baseline definition

We used data from the Stockholm CREAtinine Measurements (SCREAM) project, a healthcare utilization cohort from Stockholm, Sweden.[18] SCREAM included all Stockholm

residents aged ≥18 years who had a measurement of serum creatinine from in- or outpatient care between 2006 and 2011. SCREAM includes data from about 1.3 million adults, corresponding to 68% of the population of the region for that period. Information on demographics, disease history, vital status, pharmacy-dispensed medication, and healthcare use was obtained by linking to regional and national administrative databases.[18] All subjects with new-onset AF from January 2007 to December 2012 were selected. New-onset AF was defined as the presence of ICD-10 code I48 in any diagnostic position in primary, outpatient specialist or hospital care, with no I48 code between 1997 (when ICD coding started) and 2007. Baseline was defined as the date of first occurrence of AF. Patients were censored at the end of follow-up (31 December 2012), when they moved outside the Stockholm region or died from other causes than IS. Patients with missing data on creatinine were excluded. Since this study utilized only de-identified data, it was not deemed to require informed consent. The study was approved by the regional ethical review boards and the Swedish National Board of Welfare.

5.2.2 Outcome and predictor definitions

Study outcomes were ascertained via linkage with the government-run National Population Registry, which registers all deaths without loss to follow-up, and the National Patient Registry with diagnosis codes for essentially all (>99%) hospitalizations. The study outcome was defined as hospital admission for IS (ICD-10 codes I63x, 169.3, 169.4, and 169.8 in 1st or 2nd diagnostic position) or IS as main cause of death. Estimated glomerular filtration rate [eGFR; mL/min per 1.73 m^2, calculated with the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) formula] was calculated using the most recent measurement prior to AF diagnosis (median 0.28 years). Creatinine was measured in plasma, with either an enzymatic or corrected Jaffe method (alkaline picrate reaction); both methods are traceable to isotope dilution mass spectroscopy standards. Creatinine values <25 or >1500 μmol/L were considered outliers and were discarded. Proteinuria (median 0.87 years prior to AF diagnosis) was measured by either urinary albumin-to-creatinine ratio >30, or a urine dipstick (range: negative, 1, 2, and 3; all positive values were regarded as proteinuria). Information on disease history, including previous stroke, previous bleeding, congestive heart failure, cardiovascular disease, hypertension, and diabetes, was obtained using ICD-10 codes (detailed in **Supplementary material chapter 5, Table S1)**. The overall positive predictive value of these diagnoses in the register is about 85–95%.[19] Medication use, including antihypertensive and anti-diabetic drugs, was defined by registered pharmacy dispensations in the 180 days prior to AF diagnosis; for vitamin K antagonist or direct oral anticoagulant usage, dispensations in the 120 days before or up to 60 days after AF diagnosis were evaluated.

5.2.3 Risk scores

Risk scores to be validated were identified from a previous systematic review.[16] Based on availability of predictors, the following risk scores were validated: AFI,[20] $CHADS_2$,[21] Modified $CHADS_2$,[22] CHA_2DS_2-VASc,[23] ATRIA,[24] and GARFIELD-AF.[25] Scores were validated within the designated timeframe if specified (i.e. the prediction timeframe as specified in the original article), or within the maximum follow-up of the development cohort if no timeframe was specified. We used the same predictor definitions as the original studies where possible. An overview of the included risk scores, the original predictor definitions, and those used in this validation study is presented in the **Supplementary material chapter 5, section '*Risk Scores*'**.

5.2.4 Statistical analysis

The predictive performance of the included risk scores was assessed by their discrimination and calibration abilities, stratified by CKD stage, using the estimated glomerular filtration rate (eGFR) classification of the KDIGO (Kidney Disease: Improving Global Outcomes) criteria. Normal kidney function was defined as KDIGO G1-2 (eGFR >60 mL/min/1.73 m^2), mild CKD as KDIGO G3 (eGFR 30–60 mL/min/1.73 m^2), and advanced CKD as KDIGO G4-5 (eGFR <30 mL/min/1.73 m^2). Discrimination was assessed by the concordance index (c-index or c-statistic), which reflects how well the risk score distinguishes between patients with and without the outcome of interest. The c-statistic lies between 0.5 and 1.0, which equals pure chance and perfect discrimination, respectively. In general, c-statistic <0.7 is considered poor to moderate, 0.8 is considered good, and >0.9 excellent. For logistic risk scores, an area under the receiver operating curve was calculated. For Cox models, Harrell's c-statistic was calculated. Calibration describes the agreement between the predicted and actual probabilities of the outcome. It is typically presented in a calibration plot or calibration in the large (population average observed frequency and average predicted probability). In case of ideal calibration, the slope of the calibration curve would be 1 (i.e. a 45 degree line: predicted probability equals observed probability); for calibration in the large, the average observed and predicted probabilities would be equal. When risk scores presented an event rate instead of cumulative incidence, the cumulative incidence was approximated, as done in a previous study[16] (method detailed in **Supplement material chapter 5, section '*Formulae*'**). To assess the effect of the prediction timeframe (i.e. the time between baseline and when the outcome can occur, e.g. at 1 year for the CHA_2DS_2-VASc[23] and GARFIELD-AF[25] scores, and 5 years for the Modified $CHADS_2$ score[22]) on the predictive performance, the included risk scores were sequentially validated for different timeframes at monthly intervals, and c-statistics and calibration in the large were plotted over the entire follow-up duration of SCREAM.

This analysis may provide insight into the stability of the performance of risk scores and the dependency on the prediction timeframe. We conducted this analysis since it is not uncommon for clinicians to extrapolate or interpolate the predicted risks over time and we hypothesized that both discrimination and calibration would be highest at the timeframe for which the risk score was developed.

5.2.5 Sensitivity analyses

Three sensitivity analyses were conducted. First, an analysis using a broader, composite outcome definition of IS including transient ischemic attacks (TIAs; ICD-10 codes detailed in **Supplementary material chapter 5, Table S1**). Second, an analysis stratified on anticoagulation use, which included both vitamin K antagonists and direct oral anticoagulants. Lastly, we validated the included risk scores in subgroups with smaller eGFR ranges than the KDIGO stages to further explore the effect of eGFR on score performance. To compare the calibration in the large of the different risk scores, the mean squared error (MSE) of the average predicted and observed probabilities per eGFR cut-off ($n=11$) was calculated. The MSE is the average of the differences between the predicted and observed risks. Lower values indicate a good concordance between these risks, while higher values indicate over- or underprediction, or a combination of both. The methods used to approximate the cumulative incidence and calculate the MSE are further detailed in the **Supplementary material chapter 5, section *'formulae'*.**

5.3 RESULTS

5.3.1 Demographics

Of the 1 372 425 healthcare users in Stockholm included in SCREAM, 39 260 subjects were diagnosed with AF between 2007 and 2011, of which 3256 were excluded because of missing information on eGFR, leaving a total of 36 004 subjects eligible for analysis (**Figure 5.1**). At a median follow-up of 1.88 years, a total of 3069 (8.5%) IS occurred: 1946 (7.4%) of the 26 249 patients with normal kidney function, 1018 (11.8%) of the 8625 patients with mild CKD, and 105 (9.3%) of the 1130 patients with advanced CKD. The baseline characteristics of the included subjects, together with an overview of the study demographics of the validated risk scores, are presented in **Table 5.1.**

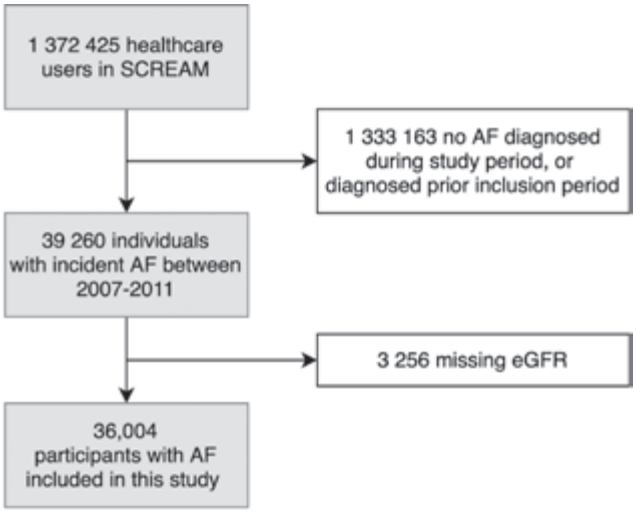


Figure 5.1 Flowchart of patient inclusion in SCREAM. AF: atrial fibrillation.

5.3.2 Discrimination

C-statistics of most risk scores were lower across worsening kidney function categories (**Table 5.2, Figure 5.2**). For the AFI score, c-statistics were 0.68 (95% confidence interval 0.67–0.69) in AF patients with normal kidney function, 0.58 (0.57–0.59) in those with mild CKD and 0.55 (0.51–0.59) in patients with advanced CKD. The c-statistics for the $CHADS_2$ were relatively stable. The Modified $CHADS_2$ score showed the highest and consistent discriminatory abilities in all kidney function groups [0.78 (0.77–0.79), 0.73 (0.71–0.74), and 0.74 (0.69–0.79), respectively]. The CHA_2DS_2-VASc score showed moderate discrimination in AF patients with normal kidney function, but poor discrimination in mild and advanced CKD. The ATRIA risk score showed good discrimination in AF patients with normal kidney function, but moderate in those with mild and advanced CKD, as did the GARFIELD-AF risk score.

Study	Validation cohort	AFI Investigators 1994[17]	Gage 2001[18]	Rietbrock 2008[19]	Lip 2010[20]	Singer 2013[21]	Fox 2017[22]
Name risk score	SCREAM	AFI	$CHADS_2$	Modified-$CHADS_2$	CHA_2DS_2-VASC	ATRIA‡	GARFIELD-AF
Total participants	36 004	4253	1 733	305 566	1 084	10 927	38 935
Total events (%)	3 069 (8.52)	106 (2.49)	94 (5.42)	19 925 (6.52)	25 (2.31)	685 (6.27)	473 (1.21)
Median follow-up	1.88y	1.2-2.3y‡‡	1.0y	2.46-2.74 y†	1y††	2.4 y	1y††
Age (SD)	74.84 (12.80)	69.4*	81	-	66	- ‡	71.0
Male sex (%)	18 891 (52.5)	2977 (70)*	728 (42)	157 202 (48.6)	642 (59.2)	- ‡	21 628 (55.5)
Race (% Caucasian)	-	3930 (92.4)*	-	-	-	-	24 157 (62.0)
CKD							
Normal eGFR (%)	26 249 (72.9)	-	-	-	-	- ‡	-
Mild CKD (%)	8 625 (24.0)	-	-	-	-	- ‡	-
Advanced (%)	1 130 (3.1)	-	-	-	-	- ‡	-
Other	-	-	-	-	-	- ‡	4038 (12.0)§§
Proteinuria (%)	2 411 (6.7)	-	-	-	-	- ‡	
Diabetes mellitus (%)	7 000 (19.4)	595 (14)*	399 (23)§	- **	187 (17.3)	- ‡	8 558 (22.0)
AF (%)	36 004 (100)	4 253 (100)*	1 733 (100)	51 807 (17.0)	1 084 (100)	10 927 (100)	38 935 (100)
Heart failure (%)	9 340 (25.9)	940 (22.1)*	970 (56)§	- **	253 (23.5)	- ‡	8 752 (22.5)
Hypertension (%)	21 966 (61.0)	1927 (45.3)*	970 (56)§	- **	729 (67.3)	- ‡	30 435 (78.2)
Previous stroke (%)	4 608 (12.8)	264 (6.2)*	433 (25)§	-	45 (4.2)	- ‡	3 030 (7.8)
Peripheral vascular disease (%)	3 069 (8.5)	404 (9.5)*	-	-	62 (5.8)	- ‡	2212 (5.7)
Ischemic heart disease	9 814 (27.3)	519 (12.2)*	-	- **	412 (38.4)	- ‡	-
Previous bleeding (%)	4 969 (13.8)	-	-	-	-	-	1 024 (2.6)
Antithrombotic drug (%)							
VKA usage (%)	14 526 (40.3)	2 113 (49.7)	0 (0)	-	0 (0)	0 (0)	16 491 (42.4)
DOAC usage (%)	217 (0.6)	-	-	-	-	-	8 804 (22.6)
Antiplatelet (%)	23 321 (64.8)	Uncertain number	529 (31)	-	802 (74.0)	-	14 084 (36.2)

Table 5.1 Baseline characteristics of the validation cohort (SCREAM) and the validated risk scores (AFI, $CHADS_2$, Modified- $CHADS_2$ CHA_2DS_2-VASC, ATRIA, GARFIELD-AF). CKD categories were defined as normal kidney function as KDIGO G1-2 (eGFR>60 ml/min/1.73m^2), mild CKD as KDIGO G3 (eGFR between 30ml/min/1,73 m^2 and 60ml/min/1,73 m^2) and advanced CKD as KDIGO G4-5 (eGFR below 30ml/min/1.73m^2). *****The AFI risk score was presented in four subgroups, with percentages instead of absolute numbers for baseline characteristics. These were estimated by calculating a weighted mean of these percentages. †Median follow-up only given per subgroup. ‡Presented their baseline characteristics per person years. §Presented as rounded percentages; absolute number estimated. **Presented as event rates; no absolute numbers. ††No information was provided on follow-up duration, but patients were censored after reaching this timepoint. ‡‡Range of mean follow-up of the included study populations. §§ Defined as CKD III-V or eGFR <60s. 0.74 (0.69–0.79), respectively]. The CHA_2DS_2-VASc score showed moderate discrimination in AF patients with normal kidney function, but poor discrimination in mild and advanced CKD. The ATRIA risk score showed good discrimination in AF patients with normal kidney function, but moderate in those with mild and advanced CKD, as did the GARFIELD-AF risk score.

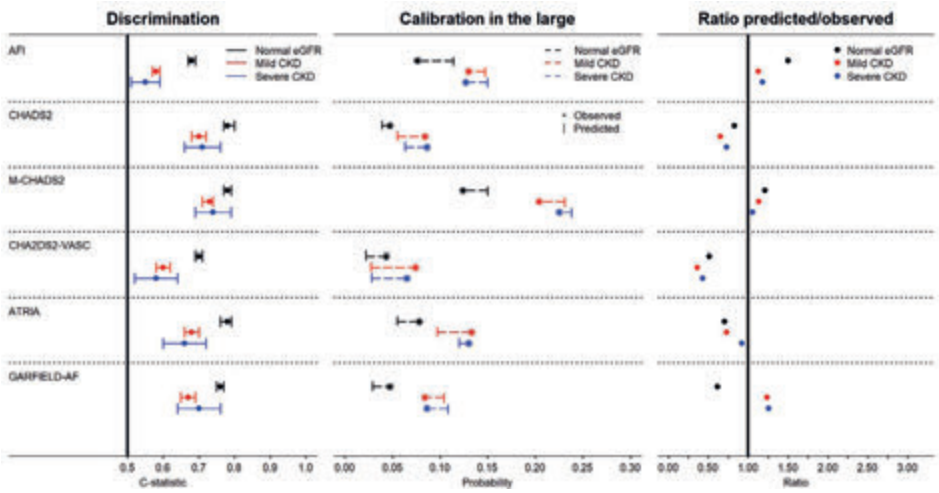


Figure 5.2 Visualization of the predictive performance, stratified by the three eGFR categories. Left: C-statistic (dot) with 95% confidence interval (bar). Middle: calibration in the large, showing the average observed (asterisk) and predicted (bar) probabilities for ischemic stroke. Left: the ratio of the predicted/observed risks – ratio's above one indicate overprediction, ratio's below underprediction. Differences in the observed risks are due to the prediction timeframe of the validated risk scores, and calculation methods (i.e. Cox or logistic).

5.3.3 Calibration

Most risk scores showed modest calibration, largely independent of kidney function (**Figures 5.2 and 5.3; Table 5.2**). For the AFI score, the calibration in the large showed overprediction in all kidney function categories. The $CHADS_2$ score underpredicted risks in the three kidney function categories. The Modified $CHADS_2$ score showed good calibration for the normal eGFR and mild CKD group, but slight overprediction in the advanced CKD group, especially so for the higher-risk patients. The CHA_2DS_2-VASc score underpredicted risks. The ATRIA score underpredicted the risk of IS. Finally, the GARFIELD-AF underpredicted the risk of IS in the normal eGFR category, but overpredicted the risks in mild and advanced CKD. The calibration plots illustrated the inaccuracy of most risk scores for patients with a high risk of IS, regardless of CKD stage, and also underlined the differences in the broadness of the prediction range (i.e. the range of possible predicted risks) (0–0.077 for CHA_2DS_2-VASC to 0.002–0.521 for GARFIELD-AF).

Risk score characteristics					Validation								
Study	**Design**	**Outcome**	**Timeframe (validated)**	**Original C-statistic**	**Normal eGFR**			**Mild CKD**			**Advanced CKD**		
					C-statistic	**Obs**	**Pred**	**C-statistic**	**Obs**	**Pred**	**C-statistic**	**Obs**	**Pred**
AFI Investigators[17] 1994, AFI	Cox	IS, TIA, SE	NS, (2.3y)	-	0.68 (0.67-0.69)	0.076	0.114	0.58 (0.57-0.59)	0.130	0.147	0.55 (0.51-0.59)	0.127	0.150
Gage[18] 2001, CHADS2	Cox	IS, TIA	NS, (1.0y)	0.82	0.78 (0.77-0.80)	0.047	0.039	0.70 (0.68-0.72)	0.084	0.055	0.71 (0.66-0.76)	0.086	0.063
Rietbrock[19] 2008, Modified-CHADS2	Cox	IS, HS	5y, (5y)	0.72	0.78 (0.77-0.79)	0.124	0.150	0.73 (0.71-0.74)	0.204	0.231	0.74 (0.69-0.79)	0.225	0.238
Lips[20] 2010, CHA2DS2-VASC	Logistic	IS, TIA, SE	1y, (1y)	0.61	0.70 (0.69-0.71)	0.043	0.022	0.60 (0.58-0.62)	0.074	0.027	0.58 (0.52-0.64)	0.065	0.028
Singer[21] 2013, ATRIA	Cox	IS, SE	NS, (2.4y)	0.73	0.78 (0.76-0.79)	0.078	0.055	0.68 (0.66-0.70)	0.133	0.097	0.66 (0.60-0.72)	0.130	0.120
Fox[22] 2017, GARFIELD-AF	Cox	IS, TIA, SE	1y, (1y)	0.69	0.76 (0.75-0.77)	0.047	0.029	0.67 (0.65-0.69)	0.084	0.104	0.70 (0.64-0.76)	0.086	0.108

Table **5.2** Overview of the predictive performance of the six included and externally validated studies. Discrimination (C-statistic) and calibration in the large (obs; observed vs pred; predicted) stratified for CKD stages. CKD categories were defined as normal kidney function as KDIGO G1-2 (eGFR>60 ml/min/1.73m^2), mild CKD as KDIGO G3 (eGFR between 30ml/min/1,73 m^2 and 60ml/min/1,73 m^2) and advanced CKD as KDIGO G4-5 (eGFR below 30ml/min/1.73m^2). Risk scores were validated at the timeframe specified in the article, or if no specification was given, at the maximum follow-up. Abbreviations: RCT: randomized controlled trial; TIA: transient ischemic attack; IS: ischemic stroke; SE: systemic embolus; HS: hemorrhagic stroke; SD: standard deviation.

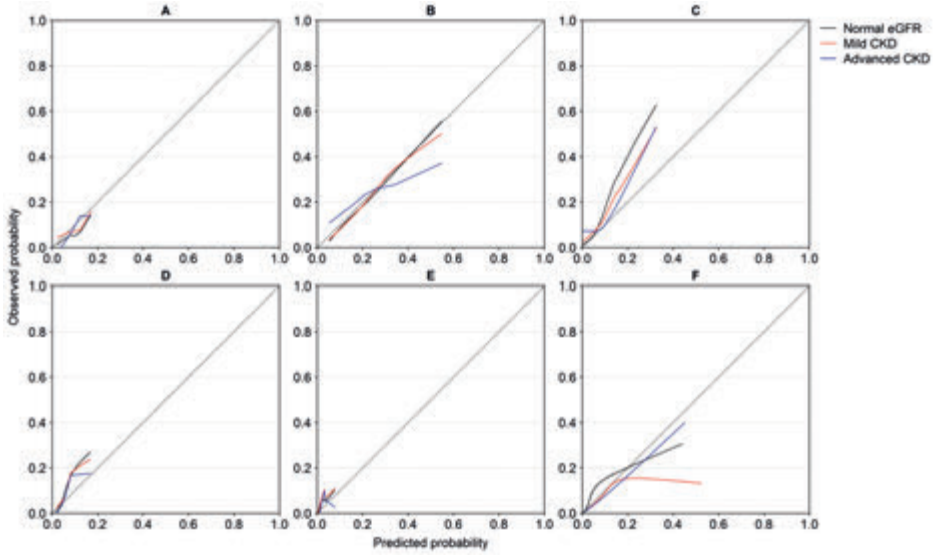


Figure 5.3 Calibration plots showing observed and predicted probabilities for ischemic stroke in patients with CKD and AF in SCREAM. Figure 5.3A: AFI; B: Modified-$CHADS_2$; C ATRIA; D: $CHADS_2$; E: CHA_2DS_2-VASc; F: GARFIELD-AF

5.3.4 Effect of the prediction timeframe on predictive performance

C-statistics were relatively stable over time, with most risk scores showing only a mild decrease in c-statistic, and subsequently stabilization (**Figure 5.4**, upper panel; stratified for CKD stages see **Supplementary material chapter 5, Figure S9**). For calibration in the large, the optimal prediction timeframe was shorter than in the development studies for $CHADS_2$ (optimal timepoint at 6 months, developed for 12 months), CHA_2DS_2-VASc (1 and 12 months, respectively), ATRIA (optimal at 17 months, validated at 29 months) and only marginally so for GARFIELD-AF (optimal at 9 months, developed for 12 months), and longer for the AFI (optimal at 49 months, validated at 28 months). The Modified $CHADS_2$ score (developed for 60 months) did not reach an optimal timepoint within 72 months (**Figure 5.4**, lower panel; stratified for CKD stages, **Supplementary material chapter 5, Figure S10**).

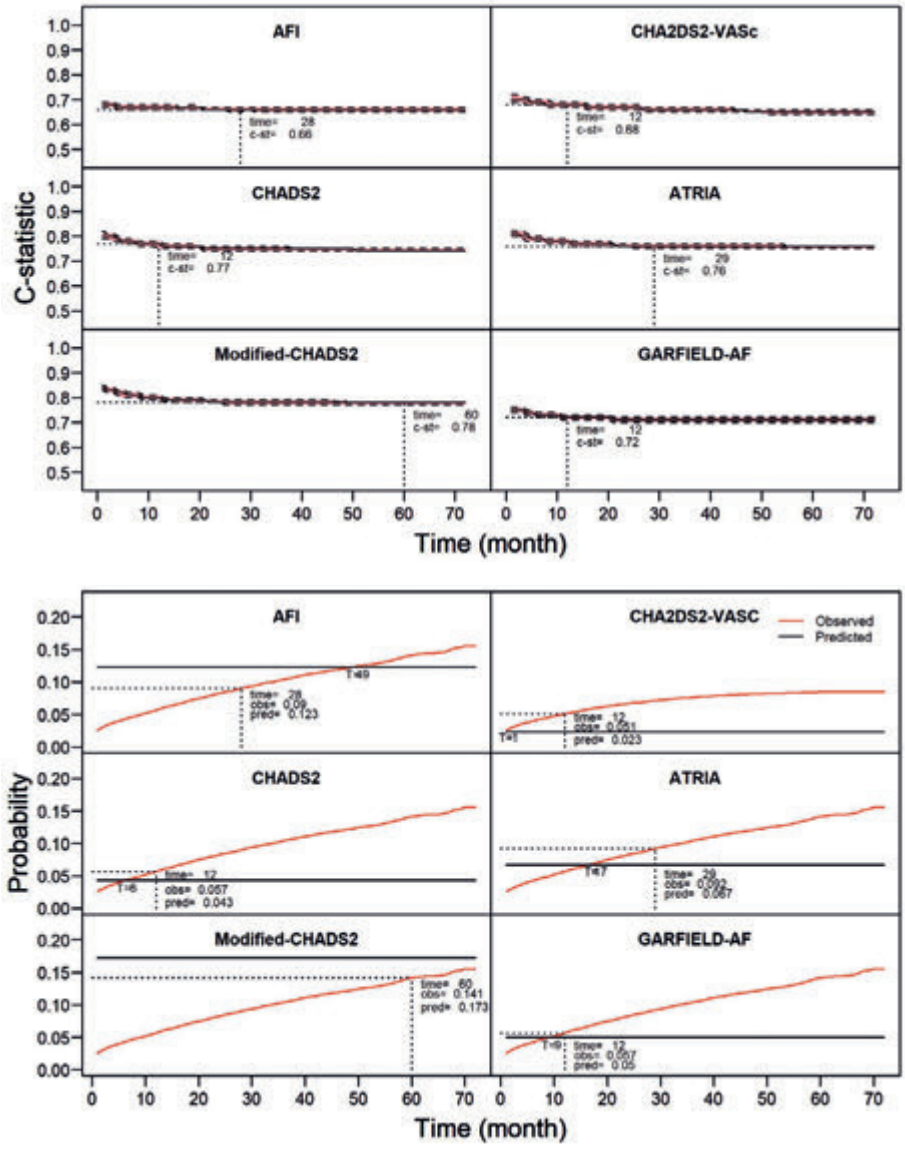


Figure 5.4 Effect of prolonging the prediction timeframe on the predictive performance in patients with AF, not stratified for CKD stage. Above the effect on discrimination (C-statistic with confidence interval); below the effect on calibration in the large. Risk scores were validated 72 times; each time prolonging the prediction timeframe with one month until the maximum follow-up of 72 months was reached. Dotted crosslines indicates the prediction timeframe for which the risk score was developed and the corresponding predictive performance, optimal calibration in the large indicated with T, followed by time in months. Stratification per CKD stage is presented in the **Supplementary material chapter 5, Figures S9-10**.

5.3.5 Sensitivity analyses

For discrimination, when validated for IS and TIA instead of IS only (sensitivity analysis 1, detailed in **Supplementary material chapter 5, Tables S2 and S3; Figure S4**), outcomes were comparable to the main analysis. Stratification by anticoagulation use (sensitivity analysis 2, **Supplementary material chapter 5, Tables S4–S7; Figure S4**) showed similar results, indicating independence from anticoagulation usage, but with broader confidence intervals due to smaller sample sizes. For most risk scores, there was a trend towards poorer discrimination in patients with lower eGFR compared with higher eGFR (sensitivity analysis 3, **Supplementary material chapter 5, Table S8, Figure S6**). The Modified $CHADS_2$ score showed consistently good discriminatory abilities, both in the main analysis and in sensitivity analyses. For calibration, the findings of the main analysis were consistent with the sensitivity analyses 1 and 2 (**Supplementary material chapter 5, Tables S3 and S5; Figures S1- 3, and S5**). The difference between the mean observed and predicted probabilities (calibration in the large) over the eGFR strata (sensitivity analysis 3, **Supplementary material chapter 5, Figure S7**) was stable, as illustrated with the low MSE values, indicating independence of the accuracy of risk scores from eGFR (**Supplementary material chapter 5, Table S9, Figure S8**). Modification of the outcome, using only ICD-10 I63x, showed similar predictive performance, though the number of events decreased to 2572 with corresponding broader confidence intervals for the c-statistics (**Supplementary material chapter 5, Section A**).

5.4 DISCUSSION

In this cohort study of 36 004 patients with AF, we externally validated six commonly used risk scores for IS. Although most risk scores showed moderate to good discrimination in patients with normal kidney function, discrimination was less accurate in moderate and advanced CKD. Calibration was largely independent of kidney function, and most risk scores either over- or underpredicted the risk of IS in one or more CKD categories. The broadness of the prediction range (i.e. the scores' ability to differentiate between low and high risks given the range of possible predicted risks) differed greatly between risk scores. The effect of the prediction timeframe influenced the predictive performance: discrimination showed an initial decrease for the shorter timeframes, but stabilized thereafter, indicating that, with regard to discrimination, risk scores can be used to predict IS on a longer or shorter prediction timeframe than designed in the original studies. For calibration, the optimal prediction timepoint differed substantially with the timepoint in the original study of most risk scores. Our results support the use of the Modified $CHADS_2$ score[22] in clinical practice as it showed good and consistent discrimination and calibration in all three kidney function categories.

Given the increasing prevalence of CKD, and the frequent use of risk scores for IS in the care of patients with CKD, there is surprisingly little information on the predictive performance in this high-risk population. Except for GARFIELD-AF and ATRIA, none of the validated risk scores included patients with CKD in their development cohorts, or CKD-specific predictors[26] in their risk score. Furthermore, external validation—the cornerstone for assessment of predictive performance in 'real-world' patients—of these risk scores is essential, but seldom performed. So far only a few studies included patients with CKD in their validation cohorts, with conflicting results: one large study on 14 264 patients with AF and eGFR >30 mL/min validated both the CHA_2DS_2-VASc and $CHADS_2$ scores showing poor discrimination (c-statistic of 0.578 and 0.575, respectively), but did not present information on calibration.[27] In another study on 307 351 patients with AF, these two risk scores performed considerably better and more in line with our results (c-statistics of 0.71 and 0.72, respectively). However, again no information on calibration was reported.[28] Finally, several studies with substantial smaller sample sizes evaluated the same risk scores, yielding comparable results, but as with the previous studies, none calculated the agreement between observed and predicted risks.[29–32] Yet, from a clinical point of view, it could be argued that this calibration, which indicates the precision of the predicted absolute risks, is clinically more important than discrimination in the setting of weighing the risks of IS and severe bleeding due to anticoagulation. This is especially relevant for patients with AF and CKD, as both the risks of IS and severe bleeding are increased.[9–11] For the clinician facing such a patient, using a risk score may seem an objective method to decide on anticoagulation therapy. However, as our study demonstrates, most of the validated risk scores for IS in this clinically relevant population either substantially over- or underpredict this risk. Although the Modified $CHADS_2$ score showed reasonable performance and would currently be the preferred risk estimation tool for patients with AF and CKD, ideally new risk scores should be developed and validated in this high-risk population. Prediction of bleeding risk appears to be equally influenced by kidney function, though data are only available for patients on dialysis.[12] This effect on predictive performance is not without consequence. Underprediction of IS risk, when weighed with bleeding risk, will result in less patients being treated with anticoagulation and consequently, an increased IS incidence, while overprediction will result in overtreatment and increased bleeding incidence. Regardless, most clinical guidelines on IS prevention recommend using the CHA_2DS_2-VASc score[8,12,13]—which showed poor predictive performance in patients with and without CKD alike. Finally, the AFI, Modified $CHADS_2$, ATRIA, and GARFIELD-AF risk scores have not been validated in patients with CKD.

Predictive performance decreased in the more clinically relevant groups of mild and advanced CKD, especially so for discrimination. Two mechanisms may have contributed to this. First, most risk scores were developed in general AF patients, and most of these studies did not include patients with CKD in their development cohort. When validating these risk scores that were developed in such heterogeneous populations in a more homogeneous population, such as patients with CKD, predictive performance—and especially discrimination—may drop.[33] While the included predictors may predict well in general AF cohorts, other more CKD-specific predictors of IS may better discriminate in this relatively homogeneous population. These include for example eGFR (which was used in ATRIA and GARFIELD-AF), proteinuria (used in ATRIA), primary kidney disease, presence of atherosclerotic vascular disease,[2] or various biomarkers (e.g. myeloperoxidase[34] or fibroblast growth factor-23,[35] amongst others[36]). Second, although we expected a comparable drop in c-statistics for the even more homogeneous patients with advanced CKD, the c-statistics of these groups were roughly equal. This may have been due to chance however: the absolute number of events in the advanced CKD group was smaller, and the level of precision consequently lower. Finally, while we ensured conformity between the predictor definitions of the original studies and our validation cohort, we deliberately validated these risk scores for the same outcome definition of IS. Indeed, most studies were developed to predict the probability of a composite outcome (e.g. CHA_2DS_2-VASc predicts a composite outcome of IS, TIA, peripheral and pulmonary embolisms). Although predictive performance might improve from validating each risk score for their specific outcome, comparability of these risk scores would then become impossible, especially when the composite outcome includes counterintuitive components, such as IS and hemorrhagic stroke. Another reason for using this outcome definition is the clinical usage: these risk scores are usually used for prediction of IS alone in the clinical setting. To test this effect, we included TIA as a composite outcome in a sensitivity analysis, which was included in most risk scores as part of the outcome. As this did not alter the results, we do not expect the effect of this outcome definition to be substantial.

5.4.1 Strengths and limitations

Our study has several strengths, but also limitations. The main strength is our large and well-defined source population, which allowed for a head-to-head comparison of multiple risk scores in well-characterized participants. Our study also provides information on calibration - the agreement between the predicted and observed risks - information that is essential for weighing IS and bleeding risk. Consistently with previous studies,[4,5] patients with more severe CKD stages in our study had a markedly increased risk of stroke.

A first limitation of our study is the large proportion of anticoagulation users in our population. Stratification for anticoagulation users and non-users yielded similar results, although discrimination was slightly poorer in anticoagulation users. Second, the cut-offs for the CKD groups might have influenced the predictive performance. Although of clinical relevance and in line with the KDIGO classification, we aimed to further explore the correlation between discrimination and kidney function. When stratified in smaller groups than the three kidney function groups, it was again shown that for most risk scores discrimination decreased with worsening of kidney function, while calibration remained relatively stable. Third, Swedish regulations do not allow the recording of ethnicity in registers, and we assumed our population to be primarily Caucasian.[37] Disparities in IS risk may be explained by ethnicity,[38] for example, blacks have a two-fold increased risk of stroke compared with non-Hispanic white adults,[39] and the predictive performance of two different scores (QRISK2 and Framingham scores) was indeed influenced by ethnicity.[40] In line with recent debates on the adequacy of the correction factor for African American ethnicity[41–43] in eGFR calculation, extrapolation of our results to other ethnicities should be done with caution. Fourth, because the prediction timeframes differed for the validated risk scores, we were unable to formally compare the predictive performance. Fifth, the use of routinely collected laboratory data may be a source of bias: for example proteinuria, a predictor used for one study (ATRIA), is not routinely measured, and measurements are performed in persons at risk. Finally, in daily clinical practice, it is not uncommon to categorize or dichotomize risk scores (e.g. CHA_2DS_2-VASc is often categorized in zero points, one point, or greater than one). Dichotomization results in loss of information[44] and our sensitivity analysis showed poor performance when validated in commonly used categories. Furthermore, most risk scores were updated many times after publication. We decided to validate the scores as intended by the authors of the original scores, instead of choosing one of the many updates or categorizations, although this may not represent the clinical use of these risk scores.

5.4.2 Implications and conclusion

Our study demonstrated moderate to poor predictive performance of various risk scores for IS in patients with AF and CKD and emphasizes how difficult this prediction is, underlining the statistical work that needs to be done in the field. For most risk scores, discriminatory abilities decreased in clinically relevant patients with mild and advanced CKD. However, calibration, which is essential for weighing the risk of IS and treatment-related bleeding, was less dependent on kidney function but still most risk scores either over- or underpredicted IS risk, or a combination of both. Prediction of IS

risk should be accurate and weighed against the risk of treatment-related bleeding. To this aim, either new scores incorporating CKD-specific predictors should be developed, or alternatively, existing and externally validated scores should be combined to increase predictive performance in this clinically relevant population, for example using ensemble modelling. As most risk scores used different prediction timeframes, this was unfeasible in our study. By conducting a head-to-head comparison of multiple scores , this study provides the clinician with information on which risk score perform well for different prediction timeframes. The Modified $CHADS_2$ score showed the best and most consistent predictive performance in all CKD stages and we suggest it is the preferred risk score to apply in clinical practice. These findings can inform the choice of risk scores in clinical practice, particularly in patients with mild to severe forms of CKD, which have not always been considered when these scores were developed.

5.5 REFERENCES

1. Feigin VL , Norrving B , Mensah GA. Global burden of stroke. Circ Res 2017;120:439–448.
2. Kumar S , Lim E , Covic A , Verhamme P , Gale CP , Camm AJ , Goldsmith D. Anticoagulation in concomitant chronic kidney disease and atrial fibrillation: JACC review topic of the week. J Am Coll Cardiol 2019;74:2204–2215.
3. Turakhia MP , Blankestijn PJ , Carrero J-J , Clase CM , Deo R , Herzog CA , Kasner SE , Passman RS , Pecoits-Filho R , Reinecke H , Shroff GR , Zareba W , Cheung M , Wheeler DC , Winkelmayer WC , Wanner C , Amann K , Banerjee D , Bansal N , Boriani G , Bunch J , Chan CT , Charytan DM , Conen D , Friedman AN , Genovesi S , Holden RM , House AA , Jadoul M , Jardine AG , Johnson DW , Jun M , Labriola L , Mark PB , McCullough PA , Nolin TD , Potpara TS , Pun PH , Ribeiro ALP , Rossignol P , Shen JI , Sood MM , Tsukamoto Y , Wang AY-M , Weir MR , Wetmore JB , Wranicz JK , Yamasaki H , Conference Participants. Chronic kidney disease and arrhythmias: conclusions from a Kidney Disease: Improving Global Outcomes (KDIGO) controversies conference. Eur Heart J 2018;39:2314–2325.
4. Cherng Y-G , Lin C-S , Shih C-C , Hsu Y-H , Yeh C-C , Hu C-J , Chen T-L , Liao C-C. Risk and outcomes in patients with chronic kidney disease or end-stage renal disease: two nationwide studies. PLoS One 2018;13:e0191155.
5. Masson P , Webster AC , Hong M , Turner R , Lindley RI , Craig JC. Chronic kidney disease and the risk of stroke: a systematic review and meta-analysis. Nephrol Dial Transplant 2015;30:1162–1169.
6. Alonso A , Lopez FL , Matsushita K , Loehr LR , Agarwal SK , Chen LY , Soliman EZ , Astor BC , Coresh J. Chronic kidney disease is associated with the incidence of atrial fibrillation: the Atherosclerosis Risk in Communities (ARIC) study. Circulation 2011;123:2946–2953.
7. Carrero JJ , Trevisan M , Sood MM , Bárány P , Xu H , Evans M , Friberg L , Szummer K. Incident atrial fibrillation and the risk of stroke in adults with chronic kidney disease: the Stockholm CREAtinine Measurements (SCREAM) project. Clin J Am Soc Nephrol 2018;13:1314–1320.
8. Kirchhof P , Benussi S , Kotecha D , Ahlsson A , Atar D , Casadei B , Castella M , Diener H-C , Heidbuchel H , Hendriks J , Hindricks G , Manolis AS , Oldgren J , Popescu BA , Schotten U , Van Putte B , Vardas P , Agewall S , Camm J , Baron Esquivias G , Budts W , Carerj S , Casselman F , Coca A , De Caterina R , Deftereos S , Dobrev D , Ferro JM , Filippatos G , Fitzsimons D , Gorenek B , Guenoun M , Hohnloser SH , Kolh P , Lip GYH , Manolis A , McMurray J , Ponikowski P , Rosenhek R , Ruschitzka F , Savelieva I , Sharma S , Suwalski P , Tamargo JL , Taylor CJ , Van Gelder IC , Voors AA , Windecker S , Zamorano JL , Zeppenfeld K. 2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS. Eur Heart J 2016;37:2893–2962.
9. Bonde AN , Lip GY , Kamper AL , Hansen PR , Lamberts M , Hommel K , Hansen ML , Gislason GH , Torp-Pedersen C , Olesen JB. Net clinical benefit of antithrombotic therapy in patients with atrial fibrillation and chronic kidney disease: a nationwide observational cohort study. J Am Coll Cardiol 2014;64:2471–2482.
10. Olesen JB , Lip GY , Kamper AL , Hommel K , Køber L , Lane DA , Lindhardsen J , Gislason GH , Torp-Pedersen C. Stroke and bleeding in atrial fibrillation with chronic kidney disease. N Engl J Med 2012;367:625–635.
11. Shah M , Avgil Tsadok M , Jackevicius CA , Essebag V , Eisenberg MJ , Rahme E , Humphries KH , Tu JV , Behlouli H , Guo H , Pilote L. Warfarin use and the risk for stroke and bleeding in patients with atrial fibrillation undergoing dialysis. Circulation 2014;129:1196–1203.

12. Camm AJ , Lip GYH , De Caterina R , Savelieva I , Atar D , Hohnloser SH , Hindricks G , Kirchhof P , Bax JJ , Baumgartner H , Ceconi C , Dean V , Deaton C , Fagard R , Funck-Brentano C , Hasdai D , Hoes A , Kirchhof P , Knuuti J , Kolh P , McDonagh T , Moulin C , Popescu BA , Reiner Ž , Sechtem U , Sirnes PA , Tendera M , Torbicki A , Vahanian A , Windecker S , Vardas P , Al-Attar N , Alfieri O , Angelini A , Blömstrom-Lundqvist C , Colonna P , De Sutter J , Ernst S , Goette A , Gorenek B , Hatala R , Heidbüchel H , Heldal M , Kristensen SD , Kolh P , Le Heuzey J-Y , Mavrakis H , Mont L , Filardi PP , Ponikowski P , Prendergast B , Rutten FH , Schotten U , Van Gelder IC , Verheugt FWA , Authors/ Task Force Members. 2012 focused update of the ESC Guidelines for the management of atrial fibrillation: an update of the 2010 ESC Guidelines for the management of atrial fibrillation—developed with the special contribution of the European Heart Rhythm Association. Europace 2012;14:1385–1413.

13. Kernan WN , Ovbiagele B , Black HR , Bravata DM , Chimowitz MI , Ezekowitz MD , Fang MC , Fisher M , Furie KL , Heck DV , Johnston SC , Kasner SE , Kittner SJ , Mitchell PH , Rich MW , Richardson D , Schwamm LH , Wilson JA. Guidelines for the prevention of stroke in patients with stroke and transient ischemic attack. Stroke 2014;45:2160–2236.

14. Hindricks G , Potpara T , Dagres N , Arbelo E , Bax JJ , Blomström-Lundqvist C , Boriani G , Castella M , Dan GA , Dilaveris PE , Fauchier L , Filippatos G , Kalman JM , La Meir M , Lane DA , Lebeau JP , Lettino M , Lip GYH , Pinto FJ , Thomas GN , Valgimigli M , Van Gelder IC , Van Putte BP , Watkins CL ; ESC Scientific Document Group. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association of Cardio-Thoracic Surgery (EACTS). Eur Heart J 2021;42:373–498.

15. Ocak G , Ramspek C , Rookmaaker MB , Blankestijn PJ , Verhaar MC , Bos WJW , Dekker FW , van Diepen M. Performance of bleeding risk scores in dialysis patients. Nephrol Dial Transplant 2019;34:1223–1231.

16. de Jong Y , Ramspek CL , van der Endt VHW , Rookmaaker MB , Blankestijn PJ , Vernooij RWM , Verhaar MC , Bos WJW , Dekker FW , Ocak G , van Diepen M. A systematic review and external validation of stroke prediction models demonstrates poor performance in dialysis patients. J Clin Epidemiol 2020;123:69–79.

17. Collins GS , Reitsma JB , Altman DG , Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 2015;350:g7594.

18. Runesson B , Gasparini A , Qureshi AR , Norin O , Evans M , Barany P , Wettermark B , Elinder CG , Carrero JJ. The Stockholm CREAtinine Measurements (SCREAM) project: protocol overview and regional representativeness. Clin Kidney J 2016;9:119–127.

19. Ludvigsson JF , Andersson E , Ekbom A , Feychting M , Kim J-L , Reuterwall C , Heurgren M , Olausson PO. External review and validation of the Swedish National Inpatient Register. BMC Public Health 2011;11:450.

20. Atrial Fibrillation Investigators. Risk factors for stroke and efficacy of antithrombotic therapy in atrial fibrillation. Analysis of pooled data from five randomized controlled trials. Arch Intern Med 1994;154:1449–1457.

21. Gage BF , Waterman AD , Shannon W , Boechler M , Rich MW , Radford MJ. Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation. JAMA 2001;285:2864–2870.

22. Rietbrock S , Heeley E , Plumb J , van Staa T. Chronic atrial fibrillation: incidence, prevalence, and prediction of stroke using the Congestive heart failure, Hypertension, Age >75, Diabetes mellitus, and prior Stroke or transient ischemic attack (CHADS2) risk stratification scheme. Am Heart J 2008;156:57–64.

23. Lip GY , Nieuwlaat R , Pisters R , Lane DA , Crijns HJ. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. Chest 2010;137:263–272.

24. Singer DE , Chang Y , Borowsky LH , Fang MC , Pomernacki NK , Udaltsova N , Reynolds K , Go AS. A new risk scheme to predict ischemic stroke and other thromboembolism in atrial fibrillation: the ATRIA study stroke risk score. J Am Heart Assoc 2013;2:e000250.

25. Fox KAA , Lucas JE , Pieper KS , Bassand JP , Camm AJ , Fitzmaurice DA , Goldhaber SZ , Goto S , Haas S , Hacke W , Kayani G , Oto A , Mantovani LG , Misselwitz F , Piccini JP , Turpie AGG , Verheugt FWA , Kakkar AK. Improved risk stratification of patients with atrial fibrillation: an integrated GARFIELD-AF tool for the prediction of mortality, stroke and bleed in patients with and without anticoagulation. BMJ Open 2017;7:e017157.

26. Ramspek CL , de Jong Y , Dekker FW , van Diepen M. Towards the best kidney failure prediction tool: a systematic review and selection aid. Nephrol Dial Transplant 2020;35:1527–1538.

27. Ravera M , Bussalino E , Fusaro M , Di Lullo L , Aucella F , Paoletti E. Systematic DOACs oral anticoagulation in patients with atrial fibrillation and chronic kidney disease: the nephrologist's perspective. J Nephrol 2020;33:483–495.

28. Friberg L , Benson L , Lip GY. Balancing stroke and bleeding risks in patients with atrial fibrillation and renal failure: the Swedish Atrial Fibrillation Cohort study. Eur Heart J 2015;36:297–306.

29. Cedeño Mora S , Goicoechea M , Torres E , Verdalles Ú , Pérez de José A , Verde E , García de Vinuesa S , Luño J. Cardiovascular risk prediction in chronic kidney disease patients. Nefrologia 2017;37:293–300.

30. Roldán V , Marín F , Manzano-Fernandez S , Fernández H , Gallego P , Valdés M , Vicente V , Lip GY. Does chronic kidney disease improve the predictive value of the CHADS2 and CHA2DS2-VASc stroke stratification risk scores for atrial fibrillation? Thromb Haemost 2013;109:956–960.

31. Bautista J , Bella A , Chaudhari A , Pekler G , Sapra KJ , Carbajal R , Baumstein D. Advanced chronic kidney disease in non-valvular atrial fibrillation: extending the utility of R2CHADS2 to patients with advanced renal failure. Clin Kidney J 2015;8:226–231.

32. Banerjee A , Fauchier L , Vourc'h P , Andres CR , Taillandier S , Halimi JM , Lip GY. Renal impairment and ischemic stroke risk assessment in patients with atrial fibrillation: the Loire Valley Atrial Fibrillation Project. J Am Coll Cardiol 2013;61:2079–2087.

33. Vergouwe Y , Moons KG , Steyerberg EW. External validity of risk models: use of benchmark values to disentangle a case-mix effect from incorrect coefficients. Am J Epidemiol 2010;172:971–980.

34. Correa S , Pena-Esparragoza JK , Scovner KM , Waikar SS , Mc Causland FR. Myeloperoxidase and the risk of CKD progression, cardiovascular disease, and death in the Chronic Renal Insufficiency Cohort (CRIC) study. Am J Kidney Dis 2020;76:32–41.

35. Scialla JJ , Xie H , Rahman M , Anderson AH , Isakova T , Ojo A , Zhang X , Nessel L , Hamano T , Grunwald JE , Raj DS , Yang W , He J , Lash JP , Go AS , Kusek JW , Feldman H , Wolf M , the Chronic Renal Insufficiency Cohort (CRIC) Study Investigators. Fibroblast growth factor-23 and cardiovascular events in CKD. J Am Soc Nephrol 2014;25:349–360.

36. Stenvinkel P , Carrero JJ , Axelsson J , Lindholm B , Heimbürger O , Massy Z. Emerging biomarkers for evaluating cardiovascular risk in the chronic kidney disease patient: how do new pieces fit into the uremic puzzle? Clin J Am Soc Nephrol 2008;3:505–521.

37. Foreign-born persons in Sweden by country of birth, age and sex. Year 2000-2015: Statistics Sweden; 2016. https://www.scb.se/en/ (16 December 2020).

38. Howard VJ , Kleindorfer DO , Judd SE , McClure LA , Safford MM , Rhodes JD , Cushman M , Moy CS , Soliman EZ , Kissela BM , Howard G. Disparities in stroke incidence contributing to disparities in stroke mortality. Ann Neurol 2011;69:619–627.

39. Roger VL , Go AS , Lloyd-Jones DM , Benjamin EJ , Berry JD , Borden WB , Bravata DM , Dai S , Ford ES , Fox CS , Fullerton HJ , Gillespie C , Hailpern SM , Heit JA , Howard VJ , Kissela BM , Kittner SJ , Lackland DT , Lichtman JH , Lisabeth LD , Makuc DM , Marcus GM , Marelli A , Matchar DB , Moy CS , Mozaffarian D , Mussolino ME , Nichol G , Paynter NP , Soliman EZ , Sorlie PD , Sotoodehnia N , Turan TN , Virani SS , Wong ND , Woo D , Turner MB ; American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics—2012 update: a report from the American Heart Association. Circulation 2012;125:e2–e220.

40. Tillin T , Hughes AD , Whincup P , Mayet J , Sattar N , McKeigue PM , Chaturvedi N , Group SS , On behalf of the SABRE Study Group. Ethnicity and prediction of cardiovascular disease: performance of QRISK2 and Framingham scores in a U.K. tri-ethnic prospective cohort study (SABRE–Southall And Brent REvisited). Heart 2014;100:60–67.

41. Diao JA , Wu GJ , Taylor HA , Tucker JK , Powe NR , Kohane IS , Manrai AK. Clinical implications of removing race from estimates of kidney function. JAMA 2021;325:184–186.

42. Powe NR. Black kidney function matters: use or misuse of race? JAMA 2020;324:737–738.

43. Eneanya ND , Yang W , Reese PP. Reconsidering the consequences of using race to estimate kidney function. JAMA 2019;322:113–114.

44. Moons KGM , Wolff RF , Riley RD , Whiting PF , Westwood M , Collins GS , Reitsma JB , Kleijnen J , Mallett S. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. Ann Intern Med 2019;170:W1–W33.

Chapter 6

APPRAISING PREDICTION RESEARCH: A GUIDE AND META-REVIEW ON BIAS AND APPLICABILITY ASSESSMENT USING THE PREDICTION MODEL RISK OF BIAS ASSESSMENT TOOL (PROBAST)

Ype de Jong
Chava L. Ramspek
Carmine Zoccali
Kitty J. Jager
Friedo W. Dekker
Merel van Diepen

Nephrology (Carlton). 2021 Dec;26(12):939-947
DOI: 10.1111/nep.13913
PMID: 34138495
PMCID: PMC9291738.

To minimize the impact of printing on the environment, the supplementary material has been uploaded. For Chapter 6, the supplement is 39 pages, and can be downloaded using the QR code.

Abstract

Over the past few years, a large number of prediction models have been published, often of poor methodological quality. Seemingly objective and straightforward, prediction models provide a risk estimate for the outcome of interest, usually based on readily available clinical information. Yet, using models of substandard methodological rigor, especially without external validation, may result in incorrect risk estimates and consequently misclassification. To assess and combat bias in prediction research the prediction model risk of bias assessment tool (PROBAST) was published in 2019. This risk of bias (ROB) tool includes four domains and 20 signaling questions highlighting methodological flaws, and provides guidance in assessing the applicability of the model. In this paper, the PROBAST will be discussed, along with an in-depth review of two commonly encountered pitfalls in prediction modelling that may induce bias: overfitting and composite endpoints. We illustrate the prevalence of potential bias in prediction models with a meta-review of 50 systematic reviews that used the PROBAST to appraise their included studies, thus including 1510 different studies on 2104 prediction models. All domains showed an unclear or high ROB; these results were markedly stable over time, highlighting the urgent need for attention on bias in prediction research. This article aims to do just that by providing (1) the clinician with tools to evaluate the (methodological) quality of a clinical prediction model, (2) the researcher working on a review with methods to appraise the included models, and (3) the researcher developing a model with suggestions to improve model quality.

6.1 INTRODUCTION AND BACKGROUND

Clinical prediction models are increasingly used for personalized medicine as these models inform on the diagnosis or the expected course of disease of individual patients. Two main groups of prediction models exist: models predicting the current presence or absence of a diagnosis (e.g., the WELLS score for screening for pulmonary embolism), and models predicting an outcome in the future (e.g., the KFRE-model for reaching end stage kidney disease in patients with chronic kidney disease). The main difference between these diagnostic- and prognostic models is the prediction timeframe, i.e. the time between the moment of prediction (i.e., baseline) and the occurrence of the outcome (respectively concurrent or in the future). After development, the predictive performance of models is typically assessed by discrimination and calibration: to what extent a model is able to differentiate between patients who reach the outcome and those who do not (discrimination), and to estimate a correct absolute risk (calibration) – concepts that are illustrated in more detail in **Box 1.**

The increased interest and use of prediction models is reflected by the abundance of newly developed prediction models. For example, we recently identified 77 models developed for ischemic stroke,[1] and 42 models predicting kidney failure in patients with chronic kidney disease (CKD).[2] This already large number of models is exceeded by far in other fields such as cardiovascular disease (estimated at nearly 800 in 2015[3]) and pulmonology (models on chronic obstructive pulmonary disease estimated at more than 450 in 2019[4]). There are likely thousands of models in other fields published in bibliographic databases, and the number of models is increasing steadily (**Figure 6.1**). Unfortunately, most of these models have come with various methodological flaws, have limited clinical uptake and have not been externally validated, meaning that the performance was assessed in new patients.[5] Although many models have found their way into everyday clinical practice, up until recently, no clear guidelines to assess a model's quality existed.

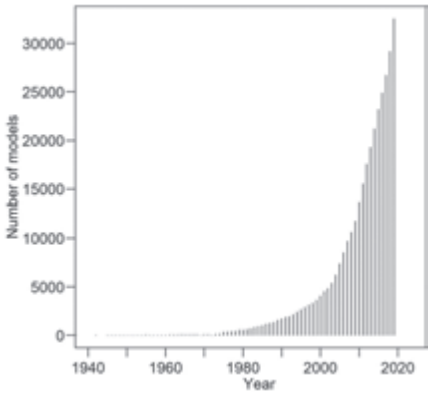


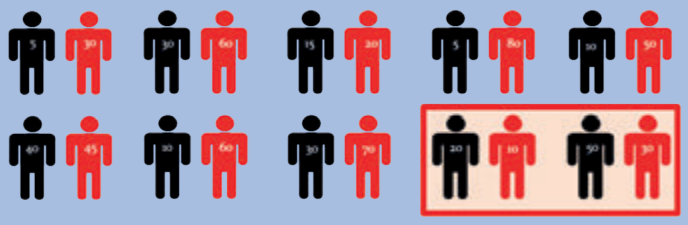
Figure 6.1 The increase in the number of prediction studies in PubMed (for the search string, see **Supplementary material chapter 6**)

Bias is usually defined as the presence of a systematic error that may affect the study's validity. However, little empirical evidence on the effects of bias in prediction research exists, and it is unclear to what extent this definition of bias in the context of etiological research is applicable to prediction. As the validity of a prediction model is tested in external validation, one way to look at bias is as a systematic difference between the model's estimated predictive performance in the development study, and the realized predictive performance in the validation study. In particular, methodological flaws in all stages of model development may result in a too optimistic estimate of predictive performance in the development cohort, which is not sustained in external settings.[6] Optimism of predictive performance is not without risk, as flawed predicted risks will result in misclassification of the outcome (i.e., poor discrimination) and inaccurate risk estimation (i.e., poor calibration). Especially when the predicted risks of two separate outcomes are compared, for example, the risk of ischemic stroke versus the risk of therapy related bleeding, misclassification is worrisome: overprediction of ischemic stroke risk will result in an increased incidence of bleeding, and vice versa.[7] Yet, the seeming simplicity and objectivity of prediction models – inserting clinical values that result in a risk of the outcome – is attractive to facilitate individualized patient care. Therefore, understanding of the potential pitfalls of prediction modelling is essential.

The aim of the present article is to provide readers with tools to appraise prediction models and assess their risk of bias (ROB) by discussing the recent publication of a ROB tool for prediction research. Next, using a meta-review approach, we will illustrate the prevalence of potential bias, and finally, we provide clear examples and illustrations of commonly encountered mistakes.

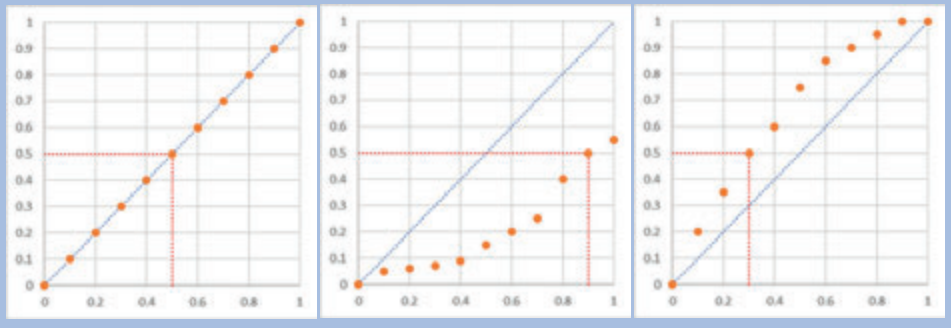
BOX 1. Discrimination and calibration

Discrimination. Describes the models' ability to discriminate between events and nonevents (logistic models) or time-to-event (Cox proportional hazard). It is typically evaluated with the area under the receiver operating curve (AUC or AUROC for logistic models) or Harrel's C-statistic (for Cox proportional hazard models) for all possible pairs of nonevents and events. Below, we visualize the mechanism behind discrimination in a sample of 20 participants, consisting of 10 nonevents (black) and 10 events (red). Numbers indicate the predicted probability. The model assigned a higher probability to the events in 8 of the 10 pairs, but a lower probability in 2 of the 10 pairs (within the marked box). Thus, the C-statistic is 0.80 (if this was a logistic model).



Discrimination:
10 pairs:
Correct n=8
Incorrect n=2
C-statistic= 0.80

Calibration. Describes the relation between the observed risks within the population, and the predicted risks. Ideally, these risk would be equal in the entire range of predicted risks (from very low to very high risk patients). Typically, calibration is assessed by calibration-in-the-large, which is the average of the predicted and the average of the observed risks. Alternatively, calibration plots (below) can be constructed showing observed risk (y-axis) per decile of predicted risk (x-axis).



Perfect calibration:
Predicted = observed (0.5 vs 0.5)

Overestimation:
Predicted > observed (0.9 vs 0.5)

Underestimation:
Predicted < observed (0.3 vs 0.5)

6.2 APPRAISING PREDICTION RESEARCH: THE PROBAST

The Prediction model Risk Of Bias ASsessment Tool (PROBAST) was published in 2019[8]: it was designed as a general tool for critical appraisal of a single prediction model study, and for the use in systematic reviews of prediction models. An elaboration, discussing the different domains and signaling questions was published separately.[6] It thus aimed to serve both the clinician that considers using a prediction model, and the researcher developing a model or including models in a systematic review or meta-analysis. The PROBAST contains two main domains: ROB and applicability. The ROB domain, which consists of four subdomains (participant selection, predictor selection, outcome definition and analysis), was defined by the authors as assessing what '(…) shortcomings in study design, conduct, or analysis could lead to systematically distorted estimates of a model's predictive performance'.[6] The applicability domain addresses concerns regarding '(…) the applicability of a primary study to the review question can arise when the population, predictors, or outcomes of the study differ from those specified in the review question'.[6] It consists of three subdomains (participant selection, predictor selection, outcome definition), and although this domain was developed for systematic reviews, the topics discussed can also apply to the use of prediction models in daily clinical practice. In total, the PROBAST contains 20 signaling questions which can be scored with low, unclear or high risk of bias, which in the end results in an overall judgement of low, unclear or high risk of bias and applicability – see **Table 6.1**. Below, we will discuss all subdomains of both ROB and applicability.

6.2.1 Risk of bias: four subdomains

The first ROB subdomain, the *participants selection subdomain*, consists of two signaling questions and concerns the use of data sources and how participants were selected, with some study designs (e.g., observational cohorts or randomized controlled trials) at lower ROB, and some at higher risk (e.g., case–control studies). In case–control studies, cases are sampled and compared to a selection of controls; therefore the percentage of cases (and absolute risk of becoming a case) does not reflect the true absolute risk. In addition, the selection of participants should represent the target population: the model should be developed in a population that is similar to the population of its intended use. For example, models developed for late-stage chronic kidney disease (CKD) should be used with caution in early CKD, and vice-versa.[9]

The second ROB subdomain, the *predictors selection subdomain*, consists of three signaling questions and covers the sources of bias that may arise due to the definition and measurement of the predictors. First, the predictors should be defined and assessed in the same way for all participants (e.g., an issue if the predictor 'body weight' was self-reported for some participants and measured for others). Furthermore, predictors should be assessed without knowledge of the outcome, that is, the outcome should be blinded when predictors are measured. In prospective research, predictor blinding is usually not an issue as the outcome is unknown at the moment predictors are established (e.g., future dialysis is not known when disease history is assessed). Finally, as a model should be usable in daily practice, all predictors should be available at the time of prediction. Although this seems rather obvious, it is not uncommon to encounter models that include predictors available only after the moment of prediction, and thus not usable in a clinical setting.

The third ROB subdomain of the PROBAST is the *outcome subdomain*. It consists of six signaling questions that may point towards sources of bias in the outcome definition. First, the outcome should be determined appropriately: misclassification of the outcome will yield biased regression coefficients. Next, the outcome should be a prespecified or standard outcome, avoiding the risk of cherry-picking an outcome that yields the best model performance. Ideally, the outcome should not include predictors in any way. Though this sounds reasonable, this is often violated: for example, several kidney failure models define this outcome as an eGFR level below a certain threshold, but meanwhile included eGFR as predictor.[2] Incorporation bias, that is, inclusion of predictors in the outcome may result in overestimation of the relation between the predictor and the outcome, and thus an overly optimistic predictive performance.[6] Clarity of outcome definitions is key: objective outcomes, such as a histological biopsy proving the presence or absence of disease, or survival versus non-survival, are less susceptible to bias than outcomes that require more interpretation of data. Obviously, if no objective outcome can be used, the outcome should be defined with such clarity that replication in a validation study, or application in the clinical field is possible. Next, this outcome should be defined and determined in a similar way for all participants. Consistent outcome definitions may be problematic in the setting of multi-center studies, where centers use different methods to assess the outcome (e.g., the presence of an ischemic stroke using a CT-scan or MRI – whichever is available). If the endpoint is a composite of multiple outcomes, these individual components should be identical for all participants – a topic that will be discussed in more detail in **Example 2**. Lastly, the time interval between the predictor assessment and the outcome determination should be appropriate to capture the clinically relevant outcome.

	1. PARTICIPANTS	2. PREDICTORS	3. OUTCOME	4. ANALYSIS
	Signaling questions			
RISK OF BIAS	1.1. Were appropriate data sources used, e.g., cohort, RCT, or nested case–control study data?	2.1. Were predictors defined and assessed in a similar way for all participants?	3.1. Was the outcome determined appropriately?	4.1. Were there a reasonable number of participants with the outcome?
	1.2. Were all inclusions and exclusions of participants appropriate?	2.2. Were predictor assessments made without knowledge of outcome data?	3.2. Was a prespecified or standard outcome definition used?	4.2. Were continuous and categorical predictors handled appropriately?
	-	2.3. Are all predictors available at the time the model is intended to be used?	3.3. Were predictors excluded from the outcome definition?	4.3. Were all enrolled participants included in the analysis?
	-	-	3.4. Was the outcome defined and determined in a similar way for all participants?	4.4. Were participants with missing data handled appropriately?
	-	-	3.5. Was the outcome determined without knowledge of predictor information?	4.5. Was selection of predictors based on univariable analysis avoided?
	-	-	3.6. Was the time interval between predictor assessment and outcome determination appropriate?	4.6. Were complexities in the data (e.g., censoring, competing risks, sampling of control participants) accounted for appropriately?
	-	-	-	4.7. Were relevant model performance measures evaluated appropriately?
	-	-	-	4.8. Were model overfitting, underfitting, and optimism in model performance accounted for?
	-	-	-	4.9. Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?
APPLICABILITY	1. PARTICIPANTS	2. PREDICTORS	3. OUTCOME	4. ANALYSIS
	Included participants or setting does not match the review question	Definition, assessment, or timing of predictors does not match the review question	Its definition, timing, or determination does not match the review question	-

Table 6.1. The domains and signaling questions of the PROBAST for assessment of risk of bias and applicability. Data presented with permission of Wolff, coauthor of the PROBAST

The final ROB subdomain concerns the *analysis*. It consists of nine signaling questions which may point to flaws in the statistical methods.[6] First, as we explain in **Example 1**, the number of events in relation to overfitting should be accounted for, especially in a setting with a limited number of events and a large number of candidate predictors. Next, continuous predictors should not be dichotomized or categorized, as this will result in loss of information, which in turn may lead to risk estimates which are imprecise. Additionally, if categorization cut-off points are based on the development dataset, for example, by using methods to identify the optimal cut-off point, the model will be overfit and biased. The next signaling question concerns the enrolment of participants in the analysis; excluding participants with outliers will likely result in bias. Furthermore, missing data – which is different from selective in–or exclusion – should be dealt with appropriately, preferably using multiple imputation instead of complete-case analysis.[10] It should be noted that studies that do not mention missing data, or methods to deal with it, likely have conducted a complete-case analysis (and are thus at increased ROB) since most statistical packages automatically exclude participants with any missing information. Next, predictor selection and accounting for competing risks and censoring should be done in a correct fashion, for example, when developing a prediction model for reaching end stage kidney disease in patients with chronic kidney disease, the competing risk of death is obvious and should be accounted for.[2] To evaluate the predictive performance – both in development and validation studies – performance measures such as discrimination and calibration should be presented in the study. Finally, all regression coefficients, including the baseline risk or model intercept, should be reported to allow the model to be used or externally validated. Many studies lack reporting baseline risk or model intercept. In addition, the presented regression coefficients should be in line with the coefficients of the final model, which may not be the case if authors retained only significant predictors of the multivariable analysis in their model, but did not re-estimate the coefficients in the smaller model. Alternatively, authors may present a risk score, where coefficients are rounded, thus losing information.

6.2.2 Assessing applicability of prognostic models

In addition to ROB assessment, the PROBAST includes signaling questions to assess the applicability of existing prognostic models for systematic reviews. Though developed for the use of systematic reviews of prediction models, we believe these subdomains are also of relevance in daily clinical use or in the setting of an external validation study.

The first subdomain of the applicability section of the PROBAST, the *participant subdomain*, considers to which extent the population in the development studies matches the participants in the review – or the clinical setting. Development studies on populations

from clinical trials, for example, need consideration: are the patients included in these trials comparable to the clinical setting, or are they healthier due to strict exclusion criteria? Furthermore, it is common for individuals enrolled in trials to be more involved in their health (i.e. healthy-user bias) which, in combination with the inclusion criteria, may result in limited external validity. Finally, models may be validated in a different specific population than intended: for example, we validated models for ischemic stroke in patients with CKD and dialysis patients.[1, 7] If the clinical rationale to do so is sound, and the model performs well, applicability may not be an issue, despite the difference in development and validation settings.

Differences in the heterogeneity of the study populations may result in differences in predictive performance. The next subdomain concerns the applicability of *predictors*, focusing on the differences in definitions, assessment or timing of predictors between the development study and the clinical setting. For example, a model using laboratory values as predictor is less applicable in a primary care setting, where blood sampling is not part of standard care. Or, alternatively, if the development study uses a predictor which value is assessed using specialized methods not routinely available, implementation in a clinical setting where this predictor is assessed using routine methods will likely result in poorer predictive performance.

The final subdomain concerns the applicability of the *outcome*. Again, as with the predictor definitions, differences in outcome definition between validation studies or the outcome of interest in the clinical practice will likely influence the predictive performance. Concerns on applicability regarding the outcome may arise if the precision of the assessment methods of the outcome differs (similarly as with applicability of predictors), but composite outcomes may also result in applicability concerns (see **Example 2**).

6.2.2.1 EXAMPLE 1. In depth review of ROB due to predictor selection

The first step of model development concerns the selection of variables predicting an outcome – either predictors with a known etiological relation or without (e.g., compare three different predictors for death: grey hair, advanced age, and telomere length – all three describe the relation of ageing with death, but with different degrees of causality). ROB regarding predictors is assessed in three subdomains of the PROBAST: the predictor definitions should be clear (predictor subdomain), predictors should not be used as or be part of the outcome (outcome subdomain), and the selection of predictors should be done in an appropriate manner (analysis). We will focus on the selection of candidate predictors for inclusion in the model, and discuss overfitting as a consequence of suboptimal selection methods in more detail.

It is common to encounter development studies that selected predictors from a list of candidate predictors by means of univariable selection: predictors with a statistically significant relation with the outcome are retained and included in the multivariable model. In univariable selection, predictors are selected based on their individual relation with the outcome, whilst in multivariable selection the strength of this relation is estimated in context with the other predictors. This univariable selection method may result in falsely rejecting predictors which may not have been statistically significant univariably but might have been in a multivariable analysis, leading to poorer predictive performance. In addition, it is susceptible to singularities in the data leading to overfitting and thus optimism– which is a major problem in prediction research. Overfitting essentially describes a prediction model fitting the development data to precisely, as depicted in **Figure 6.2**. Although the model will show good predictive performance in this dataset, outside this sample, the performance will be poor, as the model does not reflect the underlying structure of the data. Overfitting can arise at many steps. We have discussed overfitting in relation to dichotomization with data-driven cut-offs and data-driven selection methods, amongst others. Another commonly encountered cause is the sample size. The number of candidate predictors from which the final predictors are selected should be in proportion to the number of events or nonevents (whichever is smaller).

Several methods have been developed to reduce the risk of overfitting. For predictor selection, one method is to preselect predictors based on clinical knowledge or literature. Alternatively, data-driven methods such as backward selection may be performed. For this iterative method, a predefined cut-off for significance is used for in- and exclusion of predictors, and the regression components are re-estimated after each elimination. For adequate sample seizes, rules of thumb have been proposed: for Cox- and logistical modelling, 10 to 20 events per candidate predictor (e.g., meaning that, if the sample size exists of 40 events and 300 nonevents, only 2–4 candidate predictors can be used) have been suggested as appropriate sample sizes. Although the technical aspects of sample size calculations in prediction research are beyond the scope of this paper, this rule of thumb – appealing because of its simplicity – has been under critique since it was proposed in the 1990s. New and more accurate methods for sample size calculation in prediction model development studies have been proposed recently.[11] Finally, shrinkage of the model's coefficients, or internal validation (i.e., statistically simulating an external validation in the development cohort) may be conducted. However, it can be argued that the most important step for assessing and recognizing bias in prediction models is external validation: testing the capabilities of the model outside the population it was developed in.

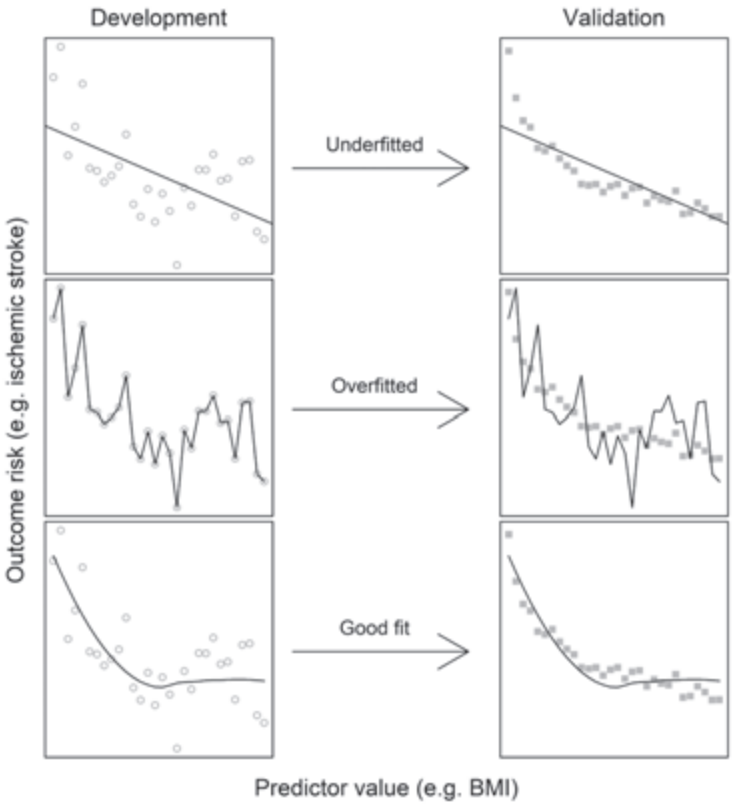


Figure 6.2 Model fitting illustrated. Different types of fit in candidate predictor selection, illustrated by two hypothetical samples of n = 30: a development cohort on the left, and a validation cohort on the right. Dots indicate the outcome risk for the predictor value (grey dots in the development cohort; grey squares in the validation cohort); the black line indicates the fitted model. BMI; Body Mass Index

6.2.2.2 EXAMPLE 2.Concerns of applicability and ROB due to composite outcomes

Composite outcomes are commonly used in prediction models: they allow developers to increase the number of events, and consequently the statistical power.[1] In the PROBAST, the authors acknowledge in two signaling questions that composite outcomes may lead to bias: as discussed above, a composite outcome should be defined beforehand and should not be adjusted based on the predictive performance. Next, the results of the individual components of the outcome should be combined in the same way for all participants. In addition, in our opinion, the clinical use should also be taken into consideration when developing a model. Take for example prediction models on the risk of ischemic stroke – a risk that should be weighed against the risk of therapy related bleeding. Whilst inclusion of

systemic embolus may be defendable from a clinical perspective, inclusion of hemorrhagic stroke in the composite outcome is odd, but encountered nonetheless.[1] Ideally, for these two entirely different outcomes, a predicted risk should be calculated and then weighed: if the risk of stroke is higher than bleeding, anticoagulation may be prescribed or vice versa, withheld. Composite outcomes combining events that require different prevention strategies (e.g., death or dialysis in chronic kidney disease,[2] death or re-transplantation after cardiac transplantation[12]) should be used with caution: if a high risk of the composite outcome is predicted, should the clinician counsel the patient for dialysis or re-transplantation, or discuss conservative therapy?

6.3 PROBAST: A META-REVIEW ON THE PREVALENCE OF BIAS

To illustrate the ROB of currently available prediction models, we conducted a systematic literature search, identifying systematic reviews that used the PROBAST for risk of bias appraisal (methods detailed in Supplementary material chapter 6, **Data S1**). One year after its publication, after removal of duplicates, we identified 151 articles that cited the PROBAST, of which 50 were systematic reviews that used the PROBAST for ROB assessment, including in total 1510 studies on 2104 models (1458 development- and 646 validation studies), see Supplementary material chapter 6, **Data S1**, Table **S1**. All of the 50 reviews presented information on ROB; 17 did not present information on applicability. Eight reviews did not present data on bias per individual study and were therefore excluded for the analysis of ROB over time. In total, of the 2104 identified models in these 50 reviews, information of ROB per domain was presented for 1039 (47%) studies (see for these models Supplementary material chapter 6, **Data S1**, **Table S2**). Overall, ROB was judged by the authors of these reviews as high: of the 1039 studies with complete information on ROB, 25% scored a high ROB in participant selection, 18% scored a high risk on predictor selection, 31% of the studies scored a high risk on the domain outcome, whilst 69% scored a high ROB in the analysis domain (**Figure 6.3**). When stratifying the ROB for the publication year of the included individual models (range 1966–2020), thus allowing visualization of trends in ROB over time, two points become clear: (1) the recent increase in prediction models, with 72% (716) of the included models published in the last 10 years and (2) though the ROB for the participant and outcome domains decreased somewhat over time, the ROB in the analysis domain remained high (**Figure 6.4**).

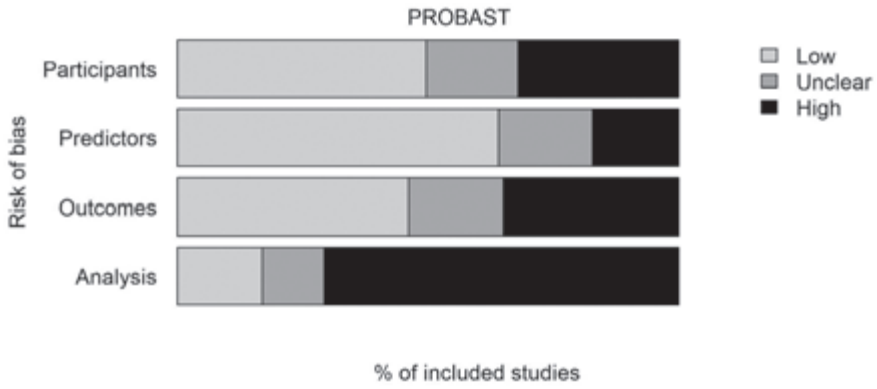


Figure 6.3 Aggregated overview of the risk of bias in 1039 prediction models with complete data (as assessed with the PROBAST in 50 systematic reviews)

6.4 PERCEIVED GAPS

Although the PROBAST, and especially the accompanying elaboration article, covers most ROB and applicability issues that may be encountered, some topics receive relatively little attention, especially regarding applicability of models in a clinical- or research setting. Though most models present information on discrimination, information on calibration is often omitted: when adapting treatment based on the absolute risk estimate, the precision of this risk estimate is essential. We suggest that models offering incomplete information on calibration should be regarded at high concern for applicability in clinic. Another topic concerns the prediction horizon: the duration of time in which the outcome could occur. The length of this prediction horizon is obviously dependent on the outcome: early warning scores predicting adverse outcomes during hospital stays will have shorter prediction horizons than risk of death due to diabetes. Regardless, the prediction horizon should be predefined (e.g., respectively at 3 days or 5 years), else clinical use or external validation is limited, as it is uncertain to what timeframe the predicted risks apply. Finally, models should present their risk estimation as an absolute risk (i.e., cumulative incidence), ideally corrected for competing risks. It is not uncommon however to encounter models presenting a hazard rate or event rate (i.e., events per person-years) instead of an absolute risk, making risk estimation and calibration cumbersome if not impossible.

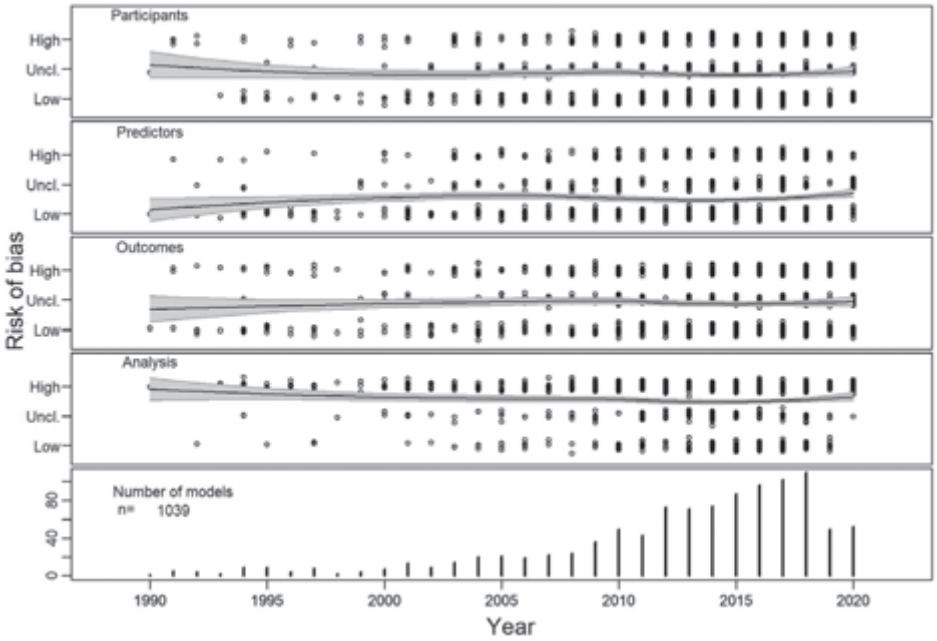


Figure 6.4 Risk of bias of 1039 prediction models extracted from 50 systematic reviews with complete data as assessed with the PROBAST, stratified per year of publication and domain (truncated at 1990). Nine hundred and eighty five models presented information on the bias domains. The trend is indicated by a fitted LOESS trendline with 95% confidence interval. For clarity, data points are jittered on the y-axis, by adding a Gaussian error with a standard deviation of 0.1

6.5 SUMMARY AND CONCLUSIONS

Prediction models are promising for individualized medicine but the overwhelming quantity and often poor quality have limited implementation in clinical practice. The PROBAST, a checklist designed to estimate ROB and assess applicability, helps the reader to determine the quality of a prediction model. This tool was well-received, as demonstrated with the already large number of systematic reviews using it just 1 year after publication. By analyzing these systematic reviews, we were able to illustrate the abundance of prediction models, and demonstrate the trends in ROB and application over time – especially so in the analysis domain. Our review of the PROBAST, together with the elaboration paper by the authors, may serve both the clinician looking to implement a model in daily practice, the researcher that aims to develop or validate a model, and the researcher conducting a systematic review on prediction models.

6.6 ACKNOWLEDGEMENT

This study was supported by a grant from the Dutch Kidney Foundation (16OKG12).

6.7 REFERENCES

1. de Jong Y, Ramspek CL, van der Endt VHW, et al. A systematic review and external validation of stroke prediction models demonstrates poor performance in dialysis patients. *Journal of clinical epidemiology* 2020;123:69-79. doi: 10.1016/j.jclinepi.2020.03.015 [published Online First: 2020/04/03]
2. Ramspek CL, de Jong Y, Dekker FW, et al. Towards the best kidney failure prediction tool: a systematic review and selection aid. *Nephrology, dialysis, transplantation* 2019 doi: 10.1093/ndt/gfz018 [published Online First: 2019/03/05]
3. Wessler BS, Lai Yh L, Kramer W, et al. Clinical Prediction Models for Cardiovascular Disease: Tufts Predictive Analytics and Comparative Effectiveness Clinical Prediction Model Database. *Circulation Cardiovascular quality and outcomes* 2015;8(4):368-75. doi: 10.1161/circoutcomes.115.001693 [published Online First: 2015/07/15]
4. Bellou V, Belbasis L, Konstantinidis AK, et al. Prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease: systematic review and critical appraisal. *BMJ-British Medical Journal* 2019;367:15. doi: 10.1136/bmj.l5358
5. Ramspek CL, Jager KJ, Dekker FW, et al. External validation of prognostic models: what, why, how, when and where? *Clinical Kidney Journal* 2020 doi: 10.1093/ckj/sfaa188
6. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Annals of internal medicine* 2019;170(1):W1-w33. doi: 10.7326/m18-1377 [published Online First: 2019/01/01]
7. de Jong Y, Fu EL, van Diepen M, et al. Validation of risk scores for ischaemic stroke in atrial fibrillation across the spectrum of kidney function. *European Heart Journal* 2021 doi: 10.1093/eurheartj/ehab059
8. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Annals of internal medicine* 2019;170(1):51-58. doi: 10.7326/m18-1376 [published Online First: 2019/01/01]
9. Ramspek CL, Evans M, Wanner C, et al. Kidney Failure Prediction Models: A Comprehensive External Validation Study in Patients with Advanced CKD. *J Am Soc Nephrol* 2021 doi: 10.1681/asn.2020071077 [published Online First: 2021/03/10]
10. de Goeij MC, van Diepen M, Jager KJ, et al. Multiple imputation: dealing with missing data. *Nephrology, dialysis, transplantation* 2013;28(10):2415-20. doi: 10.1093/ndt/gft221 [published Online First: 2013/06/05]
11. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. *Bmj* 2020;368:m441. doi: 10.1136/bmj.m441 [published Online First: 2020/03/20]
12. Aleksova N, Alba AC, Molinero VM, et al. Risk prediction models for survival after heart transplantation: A systematic review. *Am J Transplant* 2020;20(4):1137-51. doi: 10.1111/ajt.15708

6

Chapter 7

SUMMARY AND GENERAL DISCUSSION

This thesis focused on two main objectives: (1) exploring the perspectives on prognosis and person-centered care in patients with chronic kidney disease (CKD), and (2) evaluating the predictive abilities of ischemic stroke (IS) models across different levels of kidney function. Each part consists of an applied and methodological section. In **part 1**, using systematic review methods, we identified themes – common concepts, patterns or phenomena – in the qualitative studies on patients with CKD. We focused on their views on disease trajectory, prognosis and person-centered care. One of the key themes that emerged was the difference in priorities between patients and their healthcare professionals regarding outcomes of CKD. Patients emphasized the importance of being informed about the risks of these outcomes, despite the uncertainties in accurate risk estimation. We then further studied methodological aspects of reviews of qualitative studies, the method employed in **Chapter 2**. This relatively novel study design has gained traction in the last decennium, as it can be used to achieve abstraction beyond the context of the original qualitative studies.

Although patients did not explicitly mention prediction models in the qualitative review, their interest in risk prediction within the context of person-centered care was evident. Prognostic prediction models estimate the risk of an outcome of interest based on patient-specific characteristics. Prediction research in CKD traditionally emphasizes predicting the moment of dialysis initiation, or death. Though of great potential, many of these models are of methodological substandard quality, performed poorly in external validations, or lacked clinical relevance[1,2]. In **part 2** of the thesis, we identify and then validate prognostic prediction models for IS. This outcome is highly prevalent in CKD, and was one of the identified key outcomes of interest by patients with CKD in **part 1**. Yet, models predicting this future outcome receive little attention in the context of CKD, and little is known on the validity of these models within this high-risk population. We look into the predictive performance and study the stability over the full spectrum of kidney function. We conclude **part 2** with a practical guide to assess methodological quality and risk of bias (ROB) assessment of prediction models. The present chapter provides a summary of the main findings of this thesis, as well as implications on further research, both clinical and methodological.

7.1 SUMMARY OF MAIN FINDINGS

7.1.1 Summary of part 1

In **Chapter 2**, we explored the views of patients with CKD on disease trajectories, prognosis and patient centered care. For this aim, we conducted a systematic review of

qualitative studies using an inductive approach. We thematically synthesized the findings of 46 qualitative studies encompassing 1493 patients from all stages of CKD, except for those on dialysis or with a kidney transplant. While reaching end-stage kidney disease (ESKD, requiring dialysis or a kidney transplant) was commonly considered important by both patients and healthcare professionals, other outcomes that were equally important to patients were often neglected. For instance, patients mentioned a wide range of 77 symptoms that significantly impacted their daily lives. They also discussed the effects of the disease on their social life and economic productivity. Frustratingly for some patients however, these aspects of disease were not routinely discussed by healthcare professionals during outpatient care visits. This could be attributed to various barriers, including time-constrained appointments, vague or conflicting information provided, and a lack of tailoring to the patient's specific situations.

Following the diagnosis of CKD, patients experienced a period of uncertainty and social isolation, in some cases aggravated by financial burden and resentment. Perhaps because of uncertainties surrounding estimation of risks, prognosis was seldomly discussed with patients. Thus, fear of an unknown and unpredictable future, in combination with unavailable or untailored information left them in a position where they did not know what to expect, how high the risks for their prioritized outcomes were, or when certain outcomes might occur. Many patients emphasized the need to discuss prognosis and risks of outcomes relevant to them regardless of the uncertainties: they considered it better to be somewhat informed than not informed at all. A final barrier for person centered care was the feeling of being misunderstood by their healthcare professionals who either displayed a mechanistic approach to disease, instead of a holistic and person-centered approach, or had different priorities. These barriers for person centered care resulted for many patients in anxiety, frustration and resignation to regain control on their disease. This chapter illustrates the multidimensionality of patient centered care, and underlined the need for clear communication on prognosis regardless the uncertainty of estimated risks. We propose an important role for prediction models to discuss tailored risks in the patients with CKD.

In **Chapter 3**, we studied the method employed in the previous chapter, the systematic reviewing of qualitative studies. This type of review can be regarded as a 'qualitative meta-analysis': for such reviews, qualitative data and the authors' interpretations are extracted as data and then analyzed in a similar manner as original qualitative studies, thus achieving abstraction beyond the original qualitative studies. Our first aim was descriptive: how many of these reviews have been published, what type of approaches were employed, and how many studies were typically included in each review. Since the

first review of this type was published in 2007, 1.695 reviews have been published in the years following, showing an exponential trend. As with quantitative research, poorly conducted studies can lead to inappropriate findings. Hence, the second aim was to investigate the methodological quality of these studies by looking into the use of reporting quality checklists. For qualitative research, two checklists are commonly used: the COREQ (Consolidated criteria for reporting qualitative research)[3], and the ENTREQ (Enhancing transparency in reporting the synthesis of qualitative research)[4]. We analyzed these qualitative reviews both at the level of studies included in these reviews, and the review itself. The ENTREQ was adhered to more often than the COREQ (28% versus 17%), yet, the reporting quality of the qualitative studies included within these reviews improved substantially following the publication of the COREQ.

7.1.2 Summary of part 2

The second part focusses on IS risk prediction in CKD. In **Chapter 4**, we presented a systematic review and comparative head-to-head validation study of IS risk models in an incident dialysis population. Of the 77 models identified, 15 could be validated in this population (i.e. head-to-head comparison), all showing poor calibration and discriminative abilities with c-statistics ranging from 0.49 to 0.66. In this setting, the CHA_2DS_2-VASc, endorsed by the ESC guidelines, performed poorly (c-statistic 0.65 [95% CI 0.57-0.73]), and calibration was poor as well. Illustrative for the later **Chapter 6**, ROB was high for most included studies.

As discussed in the general introduction, with the pathophysiological pathways illustrated in **Figure 1.2**, risks of IS in CKD is increased in cooccurrence of atrial fibrillation (AF). The AF status of the patients included in **Chapter 4** was unknown. Thus, in **Chapter 5**, focus shifts to the remaining spectrum of kidney function in patients with AF: normal kidney function to end stage kidney disease (ESKD). Using data from the SCREAM (Stockholm CREAtinine Measurements) project, a dataset of 1.372.425 Stockholm residents with a known creatinine value, a total of 36.004 patients with new-onset AF were selected and stratified by KDIGO stage. Predictive performance of six prediction models (the AFI[5], $CHADS_2$[6], Modified $CHADS_2$[7], CHA_2DS_2-VASc[8], ATRIA[9], and GARFIELD-AF[10]) was again assessed in a head-to-head independent external validation, demonstrating a linear decline in discrimination with declining kidney function for most models. Similarly, calibration was poor, irrespective of KDIGO stage. Predictive abilities were also dependent on the prediction horizon (i.e. the time between the prediction and the occurrence of the outcome): discrimination was stable, but calibration was more dependent on this timeframe. Interestingly, ideal calibration (equal observed and predicted risks) was estimated at a different prediction horizon than the original horizon

as defined by the authors of the models (e.g. the average observed and predicted risks of the $CHADS_2$ were equal at 6 months, whereas the defined prediction horizon was 12 months). Overall, the Modified-$CHADS_2$ performed well with both a stable discrimination (c-statistics between 0.73 and 0.78) and good calibration, especially compared to the guideline-endorsed CHA_2DS_2-VASc.

External validation – and especially head-to-head validations allowing for comparison of multiple models in a relevant target population – is generally considered an essential step prior to implementation. In the previous two chapters, we demonstrated a substantial drop in predictive performance for most models compared to the development population. As discussed in the general introduction, legitimate reasons such as case mix variance may explain this. However, methodological flaws and bias probably contributed to the reduced predictive capabilities as well. Poorly developed or validated models impede implementation in clinical practice. In **Chapter 6**, we discuss the concept ROB within the setting of prediction research. We explain and expand on the recently published Prediction model Risk Of Bias ASsessment Tool (PROBAST)[11 12]. In this chapter, we aim to provide (I) the reader with tools to evaluate the methodological quality of a prediction model, (II) the researcher with methods to appraise the included models in a review, or (III) suggestions to improve model development. We illustrate the need for the PROBAST with a meta-review by plotting the ROB as assessed with the PROBAST from prediction studies included within reviews: on average, ROB was unclear or high in all domains, but especially so in the Analysis domain.

7.2 INTERPRETATION OF FINDINGS AND IMPLICATIONS FOR RESEARCH AND CLINICAL PRACTICE

The research presented in this thesis may impact both clinical practice and further research: both parts consisted of an applied section, and a methodological section, which may be of relevance to both the clinician and researcher.

7.2.1 Implications for clinical practice.

7.2.1.1 Ischemic stroke risk in CKD

In the applied sections of part 1 and 2, we explored the perspective of CKD patients on person centered care and prognostic ambiguity, and studied the predictive abilities of prediction models for IS in CKD care, respectively. At first sight, these two topics seem unrelated. However, the use of prediction models can be useful in the setting of person

centered care, by calculating a risk for an outcome specifically tailored to the patient. We demonstrated in **Chapter 2** that reaching kidney failure, necessitating dialysis was the main outcome, followed by experienced health including a total of 77 different symptoms or comorbidities of interest by CKD patients. Social and economic aspects of living with a disease were frequently mentioned as well. Risk of dying was seldomly mentioned by patients. Frustratingly for patients, apart from dialysis, neither these other possible outcomes nor the risks of these outcomes were discussed by their medical team. Yet, all of these aspects of living with CKD can be regarded as outcomes, of which, potentially, the risk could be calculated. Indeed, important steps have been taken: in a previous study by our team, we identified 42 models that predicted ESKD in CKD[2], and 16 models for the risk of death in ESKD[1]. However, despite the urgent need for prognostic clarity, a large number of prioritized outcomes by patients or by healthcare providers have not been modelled into a prediction model for CKD patients. This underlines the enormous task, but also potential for this field. In this thesis we focus on IS, one of the outcomes highlighted by patients, thus linking the applied sections of **part 1** with **part 2**.

Models for IS have been developed in patients with AF, but the validity of these models in the CKD population remains unclear. As illustrated in **Chapter 1**, **Figure 1.2**, the risk of IS in CKD is a complex interplay of shared common risk factors, disease specific risk factors, and an increased risk of atrial fibrillation (AF). With an also increased risk of bleeding, both therapy- and disease related, assessing the risk of both IS and bleeding is crucial in the CKD patient with AF. Though the relevant guidelines such as the ESC[13-15], AHA/ASA[16] acknowledge this increased risk, their approach does not differ from that of the general patient with AF. For example, since 2012, the ESC guidelines consistently advise to assess IS risk for patients with AF using the CHA_2DS_2-VASc (Congestive heart failure, Hypertension, Age >_75 years, Diabetes mellitus, Stroke, Vascular disease, Age 65-74 years, Sex category (female)). In a meta-analysis, not included in this thesis, we demonstrated the poor discriminative abilities of this endorsed score in general AF patients, yielding a very modest pooled C-statistic of 0.644 (95% confidence interval 0.635–0.653) based on 135 external validations including 3 229 267 patients with AF[17]. Yet, the predictive performance of this model in the fragile dialysis population is uncertain, as the limited number of studies that validated it yielded contradictory results, are at high ROB[18], and lacked information on calibration[19-22]. Besides the CHA_2DS_2-VASc, many more IS risk models exist, receiving markedly less attention than the CHA_2DS_2-VASc and its precursor the $CHADS_2$. The potential of identifying a 'hidden gem' – an IS risk model with good predictive capabilities in CKD and dialysis populations – is great, especially compared to the poor performance of the currently most frequently used model,

the CHA_2DS_2-VASc. This is exemplified in **Chapters 4** and **5**: though the CHA_2DS_2-VASc was the best performing model in our head-to-head validation study in dialysis patients (**Chapter 4**), it still performed modest at best with a c-statistic 0.65 (95% CI 0.57–0.73). It was thus only slightly better than the lesser known modified-$CHADS_2$ (c-statistic of 0.62 (0.56–0.68)), which was the best and most stable performing model across the spectrum of kidney function (c-statistic 0.78 (0.77–0.79) in normal eGFR, 0.73 (0.71–0.74) in mild CKD, and 0.74 (0.69–0.79) in advanced CKD not on dialysis). The discrimination of the CHA_2DS_2-VASc in normal eGFR, mild- and advanced CKD was 0.70 (0.69–0.71), 0.60 (0.58–0.62) and 0.58 (0.52–0.64), respectively.

To illustrate statistics with an example: suppose a population of 100 patients with advanced CKD, new-onset AF and an incidence of 10% for IS. Assuming there will be no competing events such as death, or loss to follow up, of these, 10 will develop an IS (an event), and 90 will not develop an IS (non-event) during the prediction horizon. Two methods are commonly used to assess discrimination, depending on the outcome: binary outcomes (for which logistic regression is usually employed) and time to event (usually assessed with Cox-proportional hazard methods) (see **textbox 7.1**).

In this example with a binary outcome, discrimination assesses whether for each pair of event and non-event, a model assigns a higher predicted risk to those that develop the outcome versus those that do not (that is, for logistical models. For Cox proportional hazard, the patient with the shorter time to event should have a higher risk. Note however, that these predicted risks are not necessarily correct; for discrimination, these risks are ranked: the highest predicted risk should be assigned to the event. The agreement between the actual, observed risk and the calculated predicted risk, is calculated with calibration measurements). In our example of 100 patients with advanced CKD and an incidence for IS of 10%, 900 possible pairs (10 events paired with 90 non-events) can be formed. The modified-$CHADS_2$ will correctly assign a higher risk to events in 666 of the 900 pairs (in **Chapter 5**, the c-statistic was 0.74 in advanced CKD; 74% of 900 is 666), whereas the CHA_2DS_2-VASc this number will be lower at 522 (c-statistic of 0.58; 58% of 900 is 522). So, 144 pairs will be misclassified by the CHA_2DS_2-VASc. Given the high incidence of AF in CKD, risk estimation using the endorsed CHA_2DS_2-VASc instead of the better performing modified-$CHADS_2$ increases the risk of misclassification and thus undertreatment with oral anticoagulants (OAC) as secondary prevention for IS.

Textbox 7.1: discrimination in logistic and Cox regression

Binary outcome, AUC or AUROC

The c-statistic in this method is the proportion of subject pairs i and j with outcome status Y (event =1, non-event =0), where the subject with the event (Y=1) had a higher predicted probability (p) than the non-event (Y=0).

$$c = \frac{\sum_{i \neq j} 1\{p_i > p_j\} 1\{Y_{i=1}\} 1\{Y_{j=0}\}}{\sum_{i \neq j} 1\{Y_{i=1}\} 1\{Y_{j=0}\}}$$

Each predicted risk can be regarded as a threshold, above and below the predicted risks can be dichotomized, and the sensitivity and specificity can be calculated. This is usually plotted in an ROC curve, a plot of sensitivity vs. 1-specificity; the c-statistics is the area under the ROC curve (AUC or AUROC).

Time to event, Harrell's c-statistic

This method uses the outcome status d, time-to-event T, and the predicted probability p for patient-pairs i and j, where $i \neq j$

$$c = \frac{\#concordant\ pairs}{\#concordant\ pairs + \#discordant\ pairs}$$

Or

$$c = \frac{\sum_{i \neq j} 1\{p_i < p_j\} 1\{T_i < T_j\} d_j}{\sum_{i \neq j} 1\{T_i > T_j\} d_j}$$

1. If both T_i and T_j are uncensored, then d can be observed. A pair is **concordant** if $p_i > p_j$ and $T_i < T_j$, and **discordant** if $p_i > p_j$ and $T_i > T_j$
2. If both T_i and T_j are censored, d will not be observed, and this pair is discarded
3. If either T_i or T_j is censored, d will only be observed for either i or j. Suppose i will develop the outcome:
   i. If $T_j < T_i$, d will not be observed for i, and this pair is discarded
   ii. If $T_j > T_i$, d will be observed for i. If $p_i > p_j$, this is pair is **concordant**, if $p_i < p_j$, its **discordant**.

The c-statistic is then computed as the number of concordant pairs divided by the total number concordant and discordant of pairs.

7.2.1.2 Weighing risks: ischemic stroke and bleeding

Besides discrimination, the agreement between the calculated predicted and actual observed risk, known as calibration, is an essential aspect of model performance. Common methods to assess calibrate include calibration in the large (the mean observed and predicted risks), the Hosmer-Lemeshow test (the sum of the predicted risks compared to the number of observed events, using chi-square statistic), and the calibration slope (the observed outcomes are regressed on the predicted risks). Alternatively, calibration can be visualized with a calibration plot, plotting the observed and predicted risks[23].

Calibration is especially relevant in a setting where risks for different outcomes are compared: e.g. the risk of IS, and the risk of therapy related bleeding. In AF without CKD, the risk of IS outweighs the risk (and consequences) of a major bleeding, and OAC should be started regardless of the estimated bleeding risk:

> "A high bleeding risk score should not lead to withholding OAC, as the net clinical benefit of OAC is even greater amongst such patients" ESC guideline 202115

The ESC guideline advises the use of the HAS-BLED[24] (Hypertension, Abnormal renal and liver function, Stroke, Bleeding, Labile INR, Elderly, Drugs or alcohol) to assess bleeding risk in order to weigh these risks. While this strong emphasis on IS risk reduction may be adequate for the average patient with AF, patients with CKD and especially advanced CKD have an increased risk of bleeding as well. In any situation where starting or withholding a treatment depends on the predicted risk, these risks should be well-calibrated, especially so when two risks are compared. If, for example, the risk of IS is overpredicted, and the risk of bleeding is underpredicted (i.e. the actual risk of IS is low, while the actual risk of bleeding is high), the patient will receive treatment that might do more harm than good.

For reasons unknown (it is not difficult to calculate), calibration is however seldomly presented in both development- and external validation studies[17 25]. Consequently, data on calibration are scarce for IS in this population, this is even more so for bleeding prediction models. In 2022, we conducted a systematic review to identify all validation studies for the ESC-endorsed CHA_2DS_2-VASc and HAS-BLED (amongst other models) in CKD and dialysis populations[18]. Besides the studies included in this thesis, few other studies have validated the CHA_2DS_2-VASc in CKD[26 27] or dialysis populations. The HAS-BLED has been validated in CKD only once[28], and twice in dialysis[19 29]. ROB was high for most of these studies, suggesting considerable methodological issues. Only one study[29] presented calibration plots; the other studies omitted this information, or provided observed event

rates stratified per risk score stratum, that were however not linked to the predicted risks. Both prediction models are risk scores (ranging from 0-9 for both the CHA_2DS_2-VASc and HAS-BLED), of which each score corresponds to an observed event rate in the development study. These observed event rates can be correlated to observed event rates in the validation studies, and can thus be regarded as a crude measure of calibration. By doing so, most studies showed either substantial over- or underprediction for both risk models[18]. These crude calibration methods should however be interpreted with caution given the high ROB of these studies.

In **Chapter 4** and **5**, we explored the calibration of IS models in dialysis and CKD populations. We evaluated 15 models in NECOSAD, large database of 1 955 incident dialysis patients. Most of these models substantially over- or underpredicted IS risk. Several models (e.g. the Modified $CHADS_2$, Framingham Heart study, GARFIELD-AF) overpredicted risks, but showed potential for recalibration by updating the intercept. The CHA_2DS_2-VASc was poorly calibrated. In the next study, in **Chapter 5**, we again demonstrated substantial over- or underprediction for five of the six models validated in normal eGFR, mild CKD and severe CKD groups. The Modified-$CHADS_2$ showed good agreement between the predicted and observed risks irrespective of kidney function. Again, the CHA_2DS_2-VASc showed poor agreement. Our two chapters highlighted another aspect of calibration: the range of predicted probabilities. Whereas the CHA_2DS_2-VASc had an absolute risk range of 0–0.077 (i.e. 0-7.7%) for a prediction horizon of one year, the range of the Modified-$CHADS_2$ was 0.03-0.54 (i.e. 3-54%), for five years. If the range of predicted probabilities is small, as for the CHA_2DS_2-VASc, the model is unable to predict those at higher risks, i.e. for the CHA_2DS_2-VASc, all patients with a risk of 0.077 and above will receive a predicted risk of 0.077. In other words, compared to other models with broader ranges, the CHA_2DS_2-VASc demonstrates a ceiling effect. Given that the Modified-$CHADS_2$ had a broader range of predicted risks, showed stable discriminative abilities, and was also well-calibrated over this entire range, the Modified-$CHADS_2$ is able to predict the risk of stroke of patients with low to high risk equally well. Based on this evidence, the main implication of these chapters is the endorsement of the Modified-$CHADS_2$ for IS risk prediction in CKD patients instead of the guideline-endorsed CHA_2DS_2-VASc.

7.2.2 Methodological considerations

7.2.2.1 Reporting guidelines

Both parts of this thesis conclude with a methodological section: in **Chapter 3**, we discuss the impact of two reporting guidelines for qualitative research, the ENTREQ and

the COREQ on reporting quality of qualitative reviews. And in **Chapter 6**, we discuss each of the 21 signaling questions (SQ) of the PROBAST, and provide an overview of the use of this ROB tool in the field. Reporting guidelines and ROB assessment tools have much in common, however as discussed below, they differ on fundamental characteristics.

The use of reporting guidelines is a widely accepted practice among scientific journals. They are used to ensure that authors are submitting articles that transparently report the necessary information, properly formatted for publication. The number of guidelines is steadily increasing, and, to quote Prof. Dr. Vandenbroucke, coauthor of the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology), '*a cynic might be forgiven for thinking that there are now so many publication guidelines that nobody can keep track of, and that they will all sink quietly into oblivion.*'[30] Indeed, the EQUATOR network (Enhancing the QUAlity and Transparency Of health Research), an initiative that keeps track of all published reporting guidelines now lists as many as 553 different reporting guidelines[31]. EQUATOR defines a reporting guideline as

> "A checklist, flow diagram, or structured text to guide authors in reporting a specific type of research, developed using explicit methodology"[31]

Different views on the merits of such guidelines exists. One view is that they help less experienced authors to avoid commonly made mistakes, recognize commonly encountered pitfalls, and provide examples on how to write their article. Reporting guidelines thus help to ensure that the content and formatting of a paper is consistent across all journals and all authors, and perhaps also improve the quality[30]. Others feel that scientists have a moral obligation to publish as clearly as possible: patients have consented to be included in their trial, and public money has been invested in their work. One of the main arguments against is that reporting guidelines can be restrictive. The guidelines may require authors to follow specific formatting and organization rules, which can limit their creativity. Lastly, reporting guidelines are regarded by some as time-consuming, as authors must ensure that their paper meets all the criteria set out in the guidelines.[32]

7.2.2.2 The COREQ in reviews: a case study of a reporting guideline as ROB tool

ROB tools are distinct from reporting guidelines in that they are designed to help authors identify and address potential sources of bias in their research. ROB tools provide a systematic approach to assessing the quality of evidence and the potential for bias in research studies. They can be used to evaluate the validity of results, identify potential sources of bias, and suggest ways to improve the quality of evidence. Bias can be

defined as a systematic error that may affect the study's validity[33]. So, where reporting guidelines look into the transparency and correctness of reporting, ROB tools focus on the methodological rigor. It is tempting to use reporting guidelines as tool to score completeness or methodological quality of a submission, thus conflating reporting guidelines with ROB tools. Of the five journals the studies in this thesis were published in, one (European Heart Journal) requires the authors to follow reporting guidelines, two 'strongly encourage the use of' (BMC Nephrology, BMC Medical Research Methodology), and two (Journal of Clinical Epidemiology, Nephrology) do not mention reporting guidelines at all. This non-uniformity in journal policies reflects the ambivalence regarding two major questions regarding reporting guidelines: (1) can reporting guidelines be used as a tool to evaluate the quality of studies, and (2) who is responsible for the correctness of the checklists[30]?

Both topics, but especially the first, were discussed in **Chapter 3**, where we studied the use of the COREQ guideline in reviews of qualitative studies. We included 1.695 qualitative reviews, published since the first review in 2007 up until 2019, and demonstrated an exponential trend in publication rates of this article type. We demonstrated three major findings of interest: first, of these reviews, 284 (14%) used the COREQ to appraise the quality of reporting of their included studies (for which, interestingly, the COREQ was not designed. It was designed for in-depth interviews and focus groups.[3]). A small subset of these reviews used the COREQ as a decision aid for inclusion in the review. This demonstrates that, in the absence of a ROB assessment tool, reporting guidelines are indeed used for quality assessment. Of the 41 guidelines included in the EQUATOR network for qualitative research, only one tool (published after our study, in 2022) is available for ROB assessment[34]. So, in this hiatus between the publication of the COREQ in 2007, and the publication of the first ROB assessment tool in 2022, authors of reviews looked for other tools to appraise their included articles.

Second, approximately half of these studies modified the COREQ for their own use, merged COREQ items with items from other reporting guidelines, or cherry-picked items from the list of 32 signaling questions (SQ). Of the 32 SQ, only 18 were used in at least 90% of the reviews. For example, adequate description of the sample (SQ 18) was checked in 98%, availability of interview guides (SQ 19) in 95%, and the use of software (SQ 27) in 94%. Author gender (SQ 4), interview characteristics (SQ 8) and description of the coding tree (SQ 25) were checked in 53%, 38%, and 54% respectively. Apparently, some SQ were regarded as more important than other by the review authors. This illustrates the unresolved debate on the benefits and drawbacks of reporting guidelines in qualitative research which, by some authors, is limiting the broadness of qualitative research as a whole[35-38]. The major implication of our study is to provide factual data on the use of the

COREQ in the field, which might facilitate this discussion.

Finally, we demonstrated a positive effect on the completeness of reporting by comparing COREQ scores of studies published prior, and post publication of the COREQ. Total COREQ scores increased from 15.51 (SE 0.31) to 17.74 (SE 0.20), and 13/32 SQ showed significant improvement, which seen in the light of the exponentially growing publication numbers, likely reflects maturation and increasing acceptance of qualitative research. Though causality cannot be inferred, the COREQ is likely to have played a role in this improvement: since its publication in 2007, it has passed the 20.000 citations. Furthermore, publication of guidelines in other fields have shown similar trends on the quality of reporting. [39-43]

7.2.2.3 ROB in prediction research: the PROBAST

Little empirical evidence on the effects of bias in prediction research exists, and it is unclear to what extent this definition of bias in the context of etiological research is applicable to prediction. The PROBAST defines ROB in this setting as:

> "(…) when shortcomings in the study design, conduct, or analysis lead to systematically distorted estimates of model predictive performance." [11 12]

The authors realize the educative aspects of their ROB tool as well:

> "Thinking about how a hypothetical prediction model study that is methodologically robust would have been designed, conducted, and analyzed helps to understand bias in study estimates of model predictive performance."

Indeed: the 33-pages long explanation and elaboration[11] is larded with practical tips, examples and 'do's and don'ts' that may serve the less experienced researcher well in recognizing potential pitfalls in his design of a development- or validation study. It is however primarily aimed at researchers conducting a systematic review or meta-analysis of prediction studies.

In line with the definition of ROB by the PROBAST authors, high ROB of development studies has been linked to poor predictive performance in their respective validation studies[44]. The tool consists of four domains (participants, predictors, outcome and analysis), and 20 SQ that can be answered as 'yes', 'probably yes', 'probably no', 'no' and 'no information'. Risk of bias per domain is judged as high if one or more SQ is answered negatively, and thus rated as high ROB. A "no" answer however does not automatically result in a high ROB domain rating; the authors leave space for the judgement of each SQ by the reviewer. Probabilistically, this makes the domains with more SQ (resp. two, three,

six and nine SQ for the four domains) more at risk for high ROB. This is illustrated in the meta-review section of **Chapter 6**, wherein we study the trends of ROB in prediction research over time, as assessed with the PROBAST: high ROB was prevalent mostly in domains 3 (31%) and 4 (69%). Trends were remarkably stable over time, a finding that was replicated in a follow-up study conducted two years following this publication[45]. Such high levels of high ROB are suggestive for a ceiling effect, meaning that it is unable to differentiate between high ROB due to one negatively answered SQ and extremely high ROB due to nine negatively answered SQs[45].

The PROBAST has been criticized as overly complex and time consuming[44]. The complexity of the tool has been acknowledged by the authors, who state that "both subject and methodological expertise"[11 12] is needed. Perhaps consequently, interrater agreement is low for many of the SQ, suggesting different interpretation of either the SQ itself, or the data to be appraised with this SQ[44 45]. Interestingly, training of graders demonstrated improved agreement[46]. Next, there is no hierarchy in the SQ: each negatively answered SQ weighs equally for the overall domain ROB. This is counterintuitive. For example, one could argue that SQ 3.5 ("Was the outcome determined without knowledge of predictor information?") will probably have a different impact on overall model performance than SQ 4.4 ("Were participants with missing data handled appropriately?"). This has led to the development of a concise version of PROBAST, which is composed of only 6 SQ that are shown to be closely associated with a high ROB[44].

In light of the shortcomings of the PROBAST, we aimed to provide additional guidance for researchers using the PROBAST, by further explaining the concepts behind each SQ in **Chapter 6**. We also discussed perceived gaps of the PROBAST. For example, the PROBAST emphasizes discrimination and has limited focus on calibration. In the setting relevant to this thesis, where IS and bleeding risks are weighted, the accuracy of these predicted risks should be evaluated in order to determine the effectiveness of the model. Next, the prediction horizon is not discussed. Ideally, developed models predict a risk for a specific timeframe, for example one year (for the CHA_2DS_2-VASc), or five years (for the Modified-$CHADS_2$). If this timeframe is not addressed in the validation study, it is likely that the model was validated on the entire follow-up period of the study, which may be longer or shorter, but is unlikely to be exactly the original prediction horizon. This notion is illustrated in a meta-analysis on all known IS prediction models and their respective validations: only 10 of the 135 validations of the CHA_2DS_2-VASc use the one-year prediction horizon; nearly all use a longer period or do not mention the timeframe at all[17]. Depending on the outcome of interest, extending the follow-up period could have an impact on the observed risk: for instance a person's chances on acquiring IS increase the

longer they are living with CKD. In **Chapter 5**, we explored the effect of the prediction horizon on both discrimination and calibration. We validated all six models 72 times, each time extending the horizon with one month. Discrimination showed a slight dip, and then stabilized. Calibration however was affected profoundly, as the number of observed events increased with the timeframe extension. Interestingly, in the graphical depiction, the lines of the predicted and observed risks crossed at different prediction horizons than those specified in the development studies, meaning that the optimal agreement between (average) predicted and observed risks may be found at different timeframes than can be expected based on the development study. This opens possibilities for clinicians aiming to predict a risk of an outcome at different timeframes, say IS risk in AF patients at 1 year, 2 years and 5 years: based on **Chapter 5**, they could use the GARFIELD-AF[10], ATRIA[9] and Modified-CHADS$_2$ for these predictions, respectively.

With the stable prevalence of high ROB (either reflecting the beforementioned ceiling effect, or actual stable trends), in an ever growing field of prediction research, the PROBAST – either the original or the proposed short form – currently remains the only tool available to assess ROB. Despite the flaws underlined above, and though complicated to use, the PROBAST remains a useful tool to assess methodological shortcomings in prediction studies. Our studies on this topic may contribute on an updated version of the PROBAST, or an altogether new ROB tool, taking the considerations outlined above into account.

7.3 IMPLICATIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

In the above sections, we have discussed the patients perspective on CKD care and prediction research, the predictive abilities of prediction models for IS in patients with CKD, and the use and merits of reporting- and ROB checklist in qualitative and prediction research. In this final paragraph, we will look forward, and discuss the potential of further research on these topics.

7.3.1 CKD: listen to the patient

Patients living with CKD are faced with uncertainties regarding *what* outcomes might occur, and *when* these outcomes can occur. These two questions – what and when – were the two main themes in **Chapter 2**, and should be addressed both in clinical CKD care and research. Regarding the first question, these outcomes, or aspects of CKD, should be interpreted in a broad sense: they include physical and mental symptoms, effects of disease on social life, and economic productivity. We demonstrated a difference

in priorities regarding outcomes of CKD between healthcare professionals and patients with CKD, as experienced by these patients. Where healthcare professionals focused mainly on the initiation of renal replacement therapies, many other outcomes were deemed equally important by patients. Patients felt being left in the dark regarding what outcomes they could expect, and many emphasized the need for holistic, patient centered care – a clear clinical implication of this chapter. The use of patient reported outcome measures (PROMs), standardized questionnaires measuring patients' perceived health, reported from the patient's perspective, may facilitate the implementation of patient centered care[47-49] and indeed is showing promising results in clinics where PROMs have been implemented[50 51].

The second question – when to expect these outcomes – was raised by many patients as well. They expressed need to discuss relevant outcomes regardless of the prognostic uncertainties: a strong call for open communication on prognosis. Though none of the included patients in **Chapter 2** specifically mentioned the use of prediction models, the added value of these models to discuss prognosis is clear. Currently, most models in CKD care are developed to predict reaching ESKD or death. Predicting other outcomes that matter to patients – e.g. those outlined in C**hapter 2**, can be of added value for patient centered care. An interesting novel development is the prediction of future PROMs, which has shown reasonable predictive performance in patients following a major trauma[52], patients receiving hip and knee replacement[53], knee arthroplasty[54], and patients with back pain[55]. To date, no models predicting PROMs within CKD care have been developed, but the potential in this field is great[56-58].

7.3.2 Qualitative research: clear reporting is essential

The role of qualitative study methods in medical research is increasing, and the number of reviews of qualitative studies is growing as well. **Chapter 3** highlighted the current use of the COREQ and ENTREQ reporting checklists in qualitative reviews, and the effect of the introduction of the COREQ on reporting clarity of original qualitative studies. Authors of reviews cherrypicked or merged items from different scoring tools and reported their appraisal incomplete or incorrectly. Exploring the quality of reporting of the qualitative studies included in those reviews, we discovered several items that were poorly reported. After the publication of the COREQ, reporting quality improved. Our study has implications for three levels: qualitative research overall, reviews of qualitative research, and original studies. First, the use of reporting guidelines in qualitative research is not undisputed[35 36 38 59 60]. By providing factual data on the use of these checklist, and the effect of the publication of the COREQ on reporting quality, these findings may facilitate

this ongoing discussion. Second, for authors writing a review, we advocate the correct use of reporting guidelines for study appraisal. Finally, authors conducting a qualitative review could cross-check their study with the items of the COREQ that scored poorly, minimizing the risk similar poor reporting.

7.3.3 Anticoagulation in CKD: weigh risk of IS with risk of bleeding

From a clinical and methodological standpoint, further research is essential to accurately assess the risk of IS and bleeding in CKD and dialysis patients. OAC is frequently denied to this vulnerable population due to fears of excessive bleeding. There are only three validation studies in CKD and dialysis populations on bleeding risk scores, but unfortunately these three studies have a high risk of bias as assessed by the PROBAST. A personalized approach where the risk of both IS and bleeding are calculated and weighed before initiating OAC, and discussed with the patient in an informed-consent setting, is much called for. To this end, an overview of all existing bleeding risk prediction models should be created, followed by external validation of these scores relevant CKD cohorts. Special focus should be given to calibration, which allows comparison of IS and bleeding risk, reducing risk of misclassification and thus incorrect treatment decisions.

7.3.4 Prediction research: do it right

Prediction models at high ROB perform poorly in validation studies[44]. As discussed above, the definition of ROB in the setting of prediction is complex. The PROBAST directs to items that may result in bias, and the examples provided in the elaboration are valuable both for the researcher conducting a prediction study, and the reviewer appraising these. Yet, there is much work to be done: as we have demonstrated, high ROB is very prevalent in prediction research, with stable trends over time. We reviewed the items of the PROBAST and discussed perceived gaps of this ROB tool, aiming to provide the researcher developing or validating a model with information to avoid common pitfalls, and the critical reader with tools to assess the ROB of the study at hand. Together with seminal papers (e.g. the PROGnosis RESearch Strategy (PROGRESS[61-64]), the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD[65 66]), and the Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK[67 68]), methodological quality of prediction research might improve.

7.3.5 Alternatives for developing new models?

There is an abundance of newly developed models. Many of those models are not subsequently externally validated, and most will never be used in clinical practice[25]. Instead of developing new models, validating existing and commonly used models in different target populations may be a viable alternative. We have demonstrated that comparative validation of multiple models in the same populations provides valuable information to identify the best performing model in this population. If needed, this model can then be recalibrated to further improve the its performance. Finally, we have explored the dependency of discrimination and calibration on the prediction timeframe, demonstrating different optimal timeframes for each model, especially with regards to calibration. Merging multiple models may be an interesting, but unexplored option to accurately predict an outcome on different timeframes.

7.4 CONCLUSION

In conclusion, this thesis explores the perspectives of patients with CKD on prognosis and person centred care, provides information on the predictive performance of models for IS in CKD and dialysis, and discusses the merits of reporting and ROB guidelines. We created a qualitative basis that can both be used in clinical care, and in research as a basis for development of PROMs. The demonstrated general poor performance and high ROB of conventional models for IS in this high-risk population underlines the work that needs to be done. Yet, based on this thesis, we endorse the modified-$CHADS_{2,}$ a point-based risk score, which showed reasonable discrimination and calibration in dialysis patients, and was stable across the spectrum of kidney function.

REFERENCES

1. Ramspek CL, Voskamp PW, van Ittersum FJ, Krediet RT, Dekker FW, van Diepen M. Prediction models for the mortality risk in chronic dialysis patients: a systematic review and independent external validation study. *Clin Epidemiol* 2017;9:451-64. doi: 10.2147/clep.S139748 [published Online First: 20170905]
2. Ramspek CL, de Jong Y, Dekker FW, van Diepen M. Towards the best kidney failure prediction tool: a systematic review and selection aid. *Nephrol Dial Transplant* 2020;35(9):1527-38. doi: 10.1093/ndt/gfz018
3. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care* 2007;19(6):349-57. doi: 10.1093/intqhc/mzm042 [published Online First: 2007/09/18]
4. Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. *BMC Med Res Methodol* 2012;12:181. doi: 10.1186/1471-2288-12-181 [published Online First: 2012/11/29]
5. Risk factors for stroke and efficacy of antithrombotic therapy in atrial fibrillation. Analysis of pooled data from five randomized controlled trials. *Arch Intern Med* 1994;154(13):1449-57. [published Online First: 1994/07/11]
6. Gage BF, Waterman AD, Shannon W, Boechler M, Rich MW, Radford MJ. Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation. *Jama* 2001;285(22):2864-70. doi: 10.1001/jama.285.22.2864 [published Online First: 2001/06/13]
7. Rietbrock S, Heeley E, Plumb J, van Staa T. Chronic atrial fibrillation: Incidence, prevalence, and prediction of stroke using the Congestive heart failure, Hypertension, Age >75, Diabetes mellitus, and prior Stroke or transient ischemic attack (CHADS2) risk stratification scheme. *Am Heart J* 2008;156(1):57-64. doi: 10.1016/j.ahj.2008.03.010 [published Online First: 2008/07/01]
8. Lip GY, Nieuwlaat R, Pisters R, Lane DA, Crijns HJ. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. *Chest* 2010;137(2):263-72. doi: 10.1378/chest.09-1584 [published Online First: 2009/09/19]
9. Singer DE, Chang Y, Borowsky LH, Fang MC, Pomernacki NK, Udaltsova N, Reynolds K, Go AS. A new risk scheme to predict ischemic stroke and other thromboembolism in atrial fibrillation: the ATRIA study stroke risk score. *J Am Heart Assoc* 2013;2(3):e000250. doi: 10.1161/jaha.113.000250 [published Online First: 2013/06/21]
10. Fox KAA, Lucas JE, Pieper KS, Bassand JP, Camm AJ, Fitzmaurice DA, Goldhaber SZ, Goto S, Haas S, Hacke W, Kayani G, Oto A, Mantovani LG, Misselwitz F, Piccini JP, Turpie AGG, Verheugt FWA, Kakkar AK. Improved risk stratification of patients with atrial fibrillation: an integrated GARFIELD-AF tool for the prediction of mortality, stroke and bleed in patients with and without anticoagulation. *BMJ Open* 2017;7(12):e017157. doi: 10.1136/bmjopen-2017-017157 [published Online First: 2017/12/24]
11. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB, Kleijnen J, Mallett S. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Ann Intern Med* 2019;170(1):W1-w33. doi: 10.7326/m18-1377
12. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB, Kleijnen J, Mallett S. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med* 2019;170(1):51-58. doi: 10.7326/m18-1376

13. Kirchhof P, Benussi S, Kotecha D, Ahlsson A, Atar D, Casadei B, Castella M, Diener HC, Heidbuchel H, Hendriks J, Hindricks G, Manolis AS, Oldgren J, Popescu BA, Schotten U, Van Putte B, Vardas P. 2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS. *Eur Heart J* 2016;37(38):2893-962. doi: 10.1093/eurheartj/ehw210 [published Online First: 20160827]

14. Camm AJ, Lip GY, De Caterina R, Savelieva I, Atar D, Hohnloser SH, Hindricks G, Kirchhof P. 2012 focused update of the ESC Guidelines for the management of atrial fibrillation: an update of the 2010 ESC Guidelines for the management of atrial fibrillation--developed with the special contribution of the European Heart Rhythm Association. *Europace* 2012;14(10):1385-413. doi: 10.1093/europace/eus305 [published Online First: 20120824]

15. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, Boriani G, Castella M, Dan GA, Dilaveris PE, Fauchier L, Filippatos G, Kalman JM, La Meir M, Lane DA, Lebeau JP, Lettino M, Lip GYH, Pinto FJ, Thomas GN, Valgimigli M, Van Gelder IC, Van Putte BP, Watkins CL. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. *Eur Heart J* 2021;42(5):373-498. doi: 10.1093/eurheartj/ehaa612

16. Kernan WN, Ovbiagele B, Black HR, Bravata DM, Chimowitz MI, Ezekowitz MD, Fang MC, Fisher M, Furie KL, Heck DV, Johnston SC, Kasner SE, Kittner SJ, Mitchell PH, Rich MW, Richardson D, Schwamm LH, Wilson JA. Guidelines for the prevention of stroke in patients with stroke and transient ischemic attack: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* 2014;45(7):2160-236. doi: 10.1161/str.0000000000000024 [published Online First: 20140501]

17. van der Endt VHW, Milders J, Penning de Vries BBL, Trines SA, Groenwold RHH, Dekkers OM, Trevisan M, Carrero JJ, van Diepen M, Dekker FW, de Jong Y. Comprehensive comparison of stroke risk score performance: a systematic review and meta-analysis among 6 267 728 patients with atrial fibrillation. *Europace* 2022 doi: 10.1093/europace/euac096 [published Online First: 2022/07/28]

18. Langehuijsen LFS, Derksen D, Milders J, Rotmans JI, Dekker FW, Van Diepen M, De Jong Y. Evaluating stroke and bleeding risk in chronic kidney disease and dialysis patients with atrial fibrillation: a systematic review and meta-analysis. *Submitted* 2023

19. Wang TK, Sathananthan J, Marshall M, Kerr A, Hood C. Relationships between Anticoagulation, Risk Scores and Adverse Outcomes in Dialysis Patients with Atrial Fibrillation. *Heart Lung Circ* 2016;25(3):243-9. doi: 10.1016/j.hlc.2015.08.012 [published Online First: 20150925]

20. Schamroth Pravda M, Cohen Hagai K, Topaz G, Schamroth Pravda N, Makhoul N, Shuvy M, Benchetrit S, Assali A, Pereg D. Assessment of the CHA2DS2-VASc Score in Predicting Mortality and Adverse Cardiovascular Outcomes of Patients on Hemodialysis. *Am J Nephrol* 2020;51(8):635-40. doi: 10.1159/000508836 [published Online First: 20200722]

21. Shih CJ, Ou SM, Chao PW, Kuo SC, Lee YJ, Yang CY, Tarng DC, Lin CC, Huang PH, Li SY, Chen YT. Risks of Death and Stroke in Patients Undergoing Hemodialysis With New-Onset Atrial Fibrillation: A Competing-Risk Analysis of a Nationwide Cohort. *Circulation* 2016;133(3):265-72. doi: 10.1161/circulationaha.115.018294 [published Online First: 20151217]

22. Chao TF, Liu CJ, Wang KL, Lin YJ, Chang SL, Lo LW, Hu YF, Tuan TC, Chung FP, Liao JN, Chen TJ, Lip GY, Chen SA. Incidence and prediction of ischemic stroke among atrial fibrillation patients with end-stage renal disease requiring dialysis. *Heart Rhythm* 2014;11(10):1752-9. doi: 10.1016/j.hrthm.2014.06.021 [published Online First: 20140618]

23. Crowson CS, Atkinson EJ, Therneau TM. Assessing calibration of prognostic risk scores. *Stat Methods Med Res* 2016;25(4):1692-706. doi: 10.1177/0962280213497434 [published Online First: 2013/08/03]

24. Pisters R, Lane DA, Nieuwlaat R, de Vos CB, Crijns HJ, Lip GY. A novel user-friendly score (HAS-BLED) to assess 1-year risk of major bleeding in patients with atrial fibrillation: the Euro Heart Survey. *Chest* 2010;138(5):1093-100. doi: 10.1378/chest.10-0134 [published Online First: 2010/03/20]

25. Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: what, why, how, when and where? *Clin Kidney J* 2021;14(1):49-58. doi: 10.1093/ckj/sfaa188 [published Online First: 2021/02/11]

26. Bautista J, Bella A, Chaudhari A, Pekler G, Sapra KJ, Carbajal R, Baumstein D. Advanced chronic kidney disease in non-valvular atrial fibrillation: extending the utility of R2CHADS2 to patients with advanced renal failure. *Clin Kidney J* 2015;8(2):226-31. doi: 10.1093/ckj/sfv006 [published Online First: 2015/03/31]

27. Apostolakis S, Guo Y, Lane DA, Buller H, Lip GY. Renal function and outcomes in anticoagulated patients with non-valvular atrial fibrillation: the AMADEUS trial. *Eur Heart J* 2013;34(46):3572-9. doi: 10.1093/eurheartj/eht328 [published Online First: 2013/08/24]

28. Suzuki M, Matsue Y, Nakamura R, Matsumura A, Hashimoto Y. Improvement of HAS-BLED bleeding score predictive capability by changing the definition of renal dysfunction in Japanese atrial fibrillation patients on anticoagulation therapy. *J Cardiol* 2014;64(6):482-7. doi: 10.1016/j.jjcc.2014.03.006 [published Online First: 2014/05/20]

29. Ocak G, Ramspek C, Rookmaaker MB, Blankestijn PJ, Verhaar MC, Bos WJW, Dekker FW, van Diepen M. Performance of bleeding risk scores in dialysis patients. *Nephrol Dial Transplant* 2019;34(7):1223-31. doi: 10.1093/ndt/gfy387 [published Online First: 2019/01/05]

30. Vandenbroucke JP. STREGA, STROBE, STARD, SQUIRE, MOOSE, PRISMA, GNOSIS, TREND, ORION, COREQ, QUOROM, REMARK… and CONSORT: for whom does the guideline toll? *Journal of Clinical Epidemiology* 2009;62(6):594-96. doi: https://doi.org/10.1016/j.jclinepi.2008.12.003[Available from: https://www.equator-network.org accessed 17-01-2023 2023.

32. Fuller T, Pearson M, Peters J, Anderson R. What affects authors' and editors' use of reporting guidelines? Findings from an online survey and qualitative interviews. *PLoS One* 2015;10(4):e0121585. doi: 10.1371/journal.pone.0121585 [published Online First: 2015/04/16]

33. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Savović J, Schulz KF, Weeks L, Sterne JAC. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928. doi: 10.1136/bmj.d5928

34. Ryan N, Vieira D, Gyamfi J, Ojo T, Shelley D, Ogedegbe O, Iwelunmor J, Peprah E. Development of the ASSESS tool: a comprehenSive tool to Support rEporting and critical appraiSal of qualitative, quantitative, and mixed methods implementation reSearch outcomes. *Implement Sci Commun* 2022;3(1):34. doi: 10.1186/s43058-021-00236-4 [published Online First: 2022/03/30]

35. Barbour RS. Checklists for improving rigour in qualitative research: a case of the tail wagging the dog? *Bmj* 2001;322(7294):1115-7. doi: 10.1136/bmj.322.7294.1115 [published Online First: 2001/05/05]

36. Kuper A, Lingard L, Levinson W. Critically appraising qualitative research. *Bmj* 2008;337:a1035. doi: 10.1136/bmj.a1035 [published Online First: 2008/08/09]

37. Sacks D, Baxter B, Campbell BCV, Carpenter JS, Cognard C, Dippel D, Eesa M, Fischer U, Hausegger K, Hirsch JA, Shazam Hussain M, Jansen O, Jayaraman MV, Khalessi AA, Kluck BW, Lavine S, Meyers PM, Ramee S, Rüfenacht DA, Schirmer CM, Vorwerk D. Multisociety Consensus Quality Improvement Revised Consensus Statement for Endovascular Therapy of Acute Ischemic Stroke. *Int J Stroke* 2018;13(6):612-32. doi: 10.1177/1747493018778713 [published Online First: 2018/05/23]

38. Dixon-Woods M, Shaw RL, Agarwal S, Smith JA. The problem of appraising qualitative research. *Qual Saf Health Care* 2004;13(3):223-5. doi: 10.1136/qhc.13.3.223 [published Online First: 2004/06/04]

39. Zamanipoor Najafabadi AH, Ramspek CL, Dekker FW, Heus P, Hooft L, Moons KGM, Peul WC, Collins GS, Steyerberg EW, van Diepen M. TRIPOD statement: a preliminary pre-post analysis of reporting and methods of prediction models. *BMJ Open* 2020;10(9):e041537. doi: 10.1136/bmjopen-2020-041537 [published Online First: 2020/09/20]

40. Bastuji-Garin S, Sbidian E, Gaudy-Marqueste C, Ferrat E, Roujeau JC, Richard MA, Canoui-Poitrine F. Impact of STROBE statement publication on quality of observational study reporting: interrupted time series versus before-after analysis. *PLoS One* 2013;8(8):e64733. doi: 10.1371/journal.pone.0064733 [published Online First: 2013/08/31]

41. Poorolajal J, Cheraghi Z, Irani AD, Rezaeian S. Quality of Cohort Studies Reporting Post the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement. *Epidemiol Health* 2011;33:e2011005. doi: 10.4178/epih/e2011005 [published Online First: 2011/07/01]

42. Sekula P, Mallett S, Altman DG, Sauerbrei W. Did the reporting of prognostic studies of tumour markers improve since the introduction of REMARK guideline? A comparison of reporting in published articles. *PLoS One* 2017;12(6):e0178531. doi: 10.1371/journal.pone.0178531 [published Online First: 2017/06/15]

43. Smidt N, Rutjes AW, van der Windt DA, Ostelo RW, Bossuyt PM, Reitsma JB, Bouter LM, de Vet HC. The quality of diagnostic accuracy studies since the STARD statement: has it improved? *Neurology* 2006;67(5):792-7. doi: 10.1212/01.wnl.0000238386.41398.30 [published Online First: 2006/09/13]

44. Venema E, Wessler BS, Paulus JK, Salah R, Raman G, Leung LY, Koethe BC, Nelson J, Park JG, van Klaveren D, Steyerberg EW, Kent DM. Large-scale validation of the prediction model risk of bias assessment Tool (PROBAST) using a short form: high risk of bias models show poorer discrimination. *J Clin Epidemiol* 2021;138:32-39. doi: 10.1016/j.jclinepi.2021.06.017 [published Online First: 2021/06/28]

45. Langehuijsen LFS, Janse R, Venema E, Kent DM, van Diepen M, Dekker FW, Steyerberg EW, de Jong Y. Systematic meta-review of prediction studies demonstrates decreasing trends in bias and low inter-rater agreement of the PROBAST. *Journal of Clinical Epidemiology* 01-2023;volume 159:159-73. doi: https://doi.org/10.1016/j.jclinepi.2023.04.012

46. Kaiser I, Pfahlberg AB, Mathes S, Uter W, Diehl K, Steeb T, Heppt MV, Gefeller O. Inter-Rater Agreement in Assessing Risk of Bias in Melanoma Prediction Studies Using the Prediction Model Risk of Bias Assessment Tool (PROBAST): Results from a Controlled Experiment on the Effect of Specific Rater Training. *J Clin Med* 2023;12(5) doi: 10.3390/jcm12051976 [published Online First: 20230302]

47. van der Willik EM, Milders J, Bart JAJ, Bos WJW, van Ittersum FJ, Ten Dam M, Hemmelder MH, Dekker FW, Meuleman Y. Discussing results of patient-reported outcome measures (PROMs) between patients and healthcare professionals in routine dialysis care: a qualitative study. *BMJ Open* 2022;12(11):e067044. doi: 10.1136/bmjopen-2022-067044 [published Online First: 2022/11/18]

48. van der Willik EM, Terwee CB, Bos WJW, Hemmelder MH, Jager KJ, Zoccali C, Dekker FW, Meuleman Y. Patient-reported outcome measures (PROMs): making sense of individual PROM scores and changes in PROM scores over time. *Nephrology* 2021;26(5):391-99. doi: https://doi.org/10.1111/nep.13843

49. Breckenridge K, Bekker HL, Gibbons E, van der Veer SN, Abbott D, Briançon S, Cullen R, Garneata L, Jager KJ, Lønning K, Metcalfe W, Morton RL, Murtagh FE, Prutz K, Robertson S, Rychlik I, Schon S, Sharp L, Speyer E, Tentori F, Caskey FJ. How to routinely collect data on patient-reported outcome and experience measures in renal registries in Europe: an expert consensus meeting. *Nephrol Dial Transplant* 2015;30(10):1605-14. doi: 10.1093/ndt/gfv209 [published Online First: 2015/05/20]

50. van der Willik EM, Hemmelder MH, Bart HAJ, van Ittersum FJ, Hoogendijk-van den Akker JM, Bos WJW, Dekker FW, Meuleman Y. Routinely measuring symptom burden and health-related quality of life in dialysis patients: first results from the Dutch registry of patient-reported outcome measures. *Clin Kidney J* 2021;14(6):1535-44. doi: 10.1093/ckj/sfz192 [published Online First: 2020/02/03]

51. Mittal SK, Ahern L, Flaster E, Maesaka JK, Fishbane S. Self-assessed physical and mental function of haemodialysis patients. *Nephrol Dial Transplant* 2001;16(7):1387-94. doi: 10.1093/ndt/16.7.1387 [published Online First: 2001/06/28]

52. Coats T, Bouamra O, Edwards A, Lecky F, Mirkes E, Sergeant J, Ivan T. 1750 Prediction of 6 month Trauma PROMS using in-hospital data. *Emergency Medicine Journal* 2022;39(12):A973-A73. doi: 10.1136/emermed-2022-RCEM2.20

53. Huber M, Kurz C, Leidl R. Predicting patient-reported outcomes following hip and knee replacement surgery using supervised machine learning. *BMC Medical Informatics and Decision Making* 2019;19(1):3. doi: 10.1186/s12911-018-0731-6

54. Anis HK, Strnad GJ, Klika AK, Zajichek A, Spindler KP, Barsoum WK, Higuera CA, Piuzzi NS. Developing a personalized outcome prediction tool for knee arthroplasty. *Bone Joint J* 2020;102-b(9):1183-93. doi: 10.1302/0301-620x.102b9.Bjj-2019-1642.R1 [published Online First: 2020/08/31]

55. Verma D, Jansen D, Bach K, Poel M, Mork PJ, d'Hollosy WON. Exploratory application of machine learning methods on patient reported data in the development of supervised models for predicting outcomes. *BMC Med Inform Decis Mak* 2022;22(1):227. doi: 10.1186/s12911-022-01973-9 [published Online First: 2022/09/02]

56. Nair D, Wilson FP. Patient-Reported Outcome Measures for Adults With Kidney Disease: Current Measures, Ongoing Initiatives, and Future Opportunities for Incorporation Into Patient-Centered Kidney Care. *Am J Kidney Dis* 2019;74(6):791-802. doi: 10.1053/j.ajkd.2019.05.025 [published Online First: 2019/09/08]

57. Eckardt KU, Bansal N, Coresh J, Evans M, Grams ME, Herzog CA, James MT, Heerspink HJL, Pollock CA, Stevens PE, Tamura MK, Tonelli MA, Wheeler DC, Winkelmayer WC, Cheung M, Hemmelgarn BR. Improving the prognosis of patients with severely decreased glomerular filtration rate (CKD G4+): conclusions from a Kidney Disease: Improving Global Outcomes (KDIGO) Controversies Conference. *Kidney Int* 2018;93(6):1281-92. doi: 10.1016/j.kint.2018.02.006 [published Online First: 20180412]

58. Couchoud C, Hemmelgarn B, Kotanko P, Germain MJ, Moranne O, Davison SN. Supportive Care: Time to Change Our Prognostic Tools and Their Use in CKD. *Clin J Am Soc Nephrol* 2016;11(10):1892-901. doi: 10.2215/cjn.12631115 [published Online First: 20160810]

59. Reynolds J, Kizito J, Ezumah N, Mangesho P, Allen E, Chandler C. Quality assurance of qualitative research: a review of the discourse. *Health Res Policy Syst* 2011;9:43. doi: 10.1186/1478-4505-9-43 [published Online First: 2011/12/21]

60. Buus N, Agdal R. Can the use of reporting guidelines in peer-review damage the quality and contribution of qualitative health care research? *Int J Nurs Stud* 2013;50(10):1289-91. doi: 10.1016/j.ijnurstu.2013.02.012 [published Online First: 2013/03/20]

61. Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, Briggs A, Udumyan R, Moons KGM, Steyerberg EW, Roberts I, Schroter S, Altman DG, Riley RD. Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes. *BMJ : British Medical Journal* 2013;346:e5595. doi: 10.1136/bmj.e5595

62. Hingorani AD, Windt DAvd, Riley RD, Abrams K, Moons KGM, Steyerberg EW, Schroter S, Sauerbrei W, Altman DG, Hemingway H. Prognosis research strategy (PROGRESS) 4: Stratified medicine research. *BMJ : British Medical Journal* 2013;346:e5793. doi: 10.1136/bmj.e5793

63. Riley RD, Hayden JA, Steyerberg EW, Moons KGM, Abrams K, Kyzas PA, Malats N, Briggs A, Schroter S, Altman DG, Hemingway H, for the PG. Prognosis Research Strategy (PROGRESS) 2: Prognostic Factor Research. *PLOS Medicine* 2013;10(2):e1001380. doi: 10.1371/journal.pmed.1001380

64. Steyerberg EW, Moons KGM, van der Windt DA, Hayden JA, Perel P, Schroter S, Riley RD, Hemingway H, Altman DG, for the PG. Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. *PLOS Medicine* 2013;10(2):e1001381. doi: 10.1371/journal.pmed.1001381

65. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD Statement. *British Journal of Surgery* 2015;102(3):148-58. doi: 10.1002/bjs.9736

66. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162(1):W1-73. doi: 10.7326/m14-0698

67. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM, for the Statistics Subcommittee of the NCIEWGoCD. REporting recommendations for tumour MARKer prognostic studies (REMARK). *British Journal of Cancer* 2005;93(4):387-91. doi: 10.1038/sj.bjc.6602678

68. Sauerbrei W, Taube SE, McShane LM, Cavenagh MM, Altman DG. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): An Abridged Explanation and Elaboration. *JNCI: Journal of the National Cancer Institute* 2018;110(8):803-11. doi: 10.1093/jnci/djy088

7

APPENDICES

APPENDIX 1: SUMMARY IN DUTCH

Nederlandse samenvatting

A1.1 Achtergrond en doelstelling

Gezonde nieren zijn essentieel voor het handhaven van homeostase in tal van processen: ze reguleren de elektrolytenconcentratie en zuur-base balans, scheiden toxines en afvalstoffen uit via de urine, spelen een rol in de mineralisatie van botten en reguleren de bloeddruk en extracellulaire vochtbalans. Schade aan deze vitale organen, zowel acuut als chronisch, leidt daardoor tot verstoring van deze processen en een grote diversiteit aan mogelijke uitkomsten. Dit laatste hoofdstuk biedt een samenvatting van de belangrijkste bevindingen gepubliceerd in deze thesis, en beschrijft de implicaties voor verder onderzoek, zowel klinisch als methodologisch.

In het eerste deel van deze thesis onderzoeken we wat het betekent voor mensen om – althans in de ogen van zorgverleners – plots patiënt worden, nadat ze gediagnosticeerd zijn met chronische nierinsuff i ciëntie (chronic kidney disease; CKD), een ziekte die doorgaans asymptomatisch is, en bij de algemene bevolking relatief onbekend is. Wat houdt hen bezig na deze diagnose? Hoe gaan zij om met de realiteit van een vaak progressieve ziekte? Maar ook: hoe ervaren zij de zorg? Zijn er in hun ogen knelpunten, en waar is ruimte voor verbetering? CKD is een heterogene groep van nierziekten, en ontstaat vaker bij personen met diabetes mellitus, hypertensie, hart- en vaatziekten en auto-immuunziekten. Het treft 9,1% (95% CI 8,5-9,8) van de wereldbevolking en neemt wereldwijd in prevalentie toe. Afhankelijk van deze onderliggende ziekte kan de nierschade herstellen (acute nierinsufficiëntie; AKI), stabiliseren (CKD) of ontwikkelen tot nierfalen (end stage kidney disease; ESKD). In tegenstelling tot de grotendeels symptoomvrije periode van vroegere stadia van CKD, kunnen er in dit laatste stadium talrijke symptomen optreden. Deze kunnen gerelateerd zijn aan de onderliggende ziekte, aan de CKD zelf, of aan een samenspel hiervan. Onzekerheid over de kansen op deze complicaties, veroorzaakt bijvoorbeeld door onduidelijkheid over de etiologie van de CKD, of de effectiviteit van therapie hiervoor, leidt er vaak toe dat in de spreekkamer minder over prognose wordt gesproken dan patiënten zouden willen.

Centraal in het tweede deel van deze thesis staat een veelvoorkomende complicatie van CKD: het ontwikkelen van een ischemische beroerte (ischemic stroke; IS). We onderzoeken of, met behulp van predictiemodellen – een wiskundige vergelijking, ontwikkeld op basis van patiëntgegevens – de clinicus het risico op deze uitkomst in de spreekkamer bespreekbaar kan maken. Binnen de nefrologie ligt de focus van deze modellen voornamelijk op het voorspellen van ESKD bij patiënten met CKD, en het risico op sterfte bij patiënten

met ESKD. Het implementeren van deze predictiemodellen in de routinematige zorg wordt belemmerd door methodologische problemen, een hoog risico op bias (ROB) en het ontbreken van informatie over de nauwkeurigheid van de voorspellingen in deze populatie. Dit laatste wordt bepaald in een zgn. externe validatie, een essentiële benchmarktest voor implementatie waarbij wordt onderzocht hoe goed het model functioneert in de doelpopulatie. Twee maatstaven zijn essentieel voor voorspellend onderzoek: discriminatie, dat is de mate waarin het model kan differentiëren tussen patiënten die de uitkomst wel en niet krijgen, en kalibratie; de overeenkomst tussen de voorspelde en waargenomen risico's.

Deze thesis had twee doelen: (1) het verkennen van perspectieven met betrekking tot prognose als onderdeel van persoonsgerichte zorg bij patiënten met CKD, en (2) het evalueren van predictiemodellen voor IS bij verschillende stadia van nierfunctie. Elk deel bestaat uit een toegepast en methodologisch gedeelte. In **deel 1** identificeerden we thema's in kwalitatieve studies over CKD, waarbij we ons richtten op de visie van patiënten op ziekteverloop, prognose en persoonsgerichte zorg. Een belangrijk thema is het verschil in prioriteiten tussen patiënten en zorgverleners met betrekking tot het bespreken van mogelijke uitkomsten van CKD: patiënten benadrukten dat zij geïnformeerd willen worden over de risico's, ondanks de onzekerheden over de risicoschatting. In **deel 2** gaan we dieper in op het voorspellen van risico's, en identificeren we modellen voor het voorspellen van IS die wij vervolgens extern valideren. We keken naar de validiteit en stabiliteit van deze modellen over het volledige spectrum van nierfunctie: van patiënten met een ongestoorde nierfunctie, tot het moment na starten met dialyse. We sluiten beide delen af met respectievelijk een analyse over rapportagerichtlijnen voor kwalitatief onderzoek, en een praktische handleiding voor het beoordelen van de methodologische kwaliteit en (risk of bias; ROB).

A1.2 SAMENVATTING VAN DE BELANGRIJKSTE BEVINDINGEN

A1.2.1 Samenvatting van deel 1

In **Hoofdstuk 2** verkennen we de visies van patiënten met CKD over ziekteverloop, prognose en zorg op maat; ook wel persoonsgerichte zorg genoemd. Voor dit doel verrichtten we een systematische review van kwalitatieve studies, waarbij we de bevindingen van 46 kwalitatieve studies, met in totaal 1493 patiënten in alle stadia van CKD (uitgezonderd dialysepatiënten of patiënten met een niertransplantatie), thematisch en inductief analyseerden.

Hoewel het bereiken van ESKD, wat een keuze tussen dialyse of een niertransplantatie

noodzaakt, vaak als belangrijk werd beschouwd door zowel patiënten als zorgprofessionals, werden andere uitkomsten die even belangrijk waren voor patiënten vaak niet besproken. Patiënten noemden bijvoorbeeld een breed scala van fysieke en mentale symptomen die aanzienlijke impact hadden op hun dagelijks leven. Daarnaast benadrukten zij de effecten van de ziekte op hun sociale leven en economische productiviteit. Deze aspecten van CKD werden zelden besproken met patiënten, waardoor zij in onzekerheid verkeerden over wat ze konden verwachten en hoe hoog de risico's hierop waren. Veel patiënten benadrukten de behoefte aan het bespreken van prognose en risico's, zelfs als er onzekerheid was, omdat ze liever enigszins dan helemaal niet geïnformeerd willen worden. Patiënten die geïnformeerd werden ervoeren meer gevoel van controle over hun ziekte, dan patiënten bij wie deze gesprekken minder of niet aangingen met hun behandelaar – zij noemden vaak een conflict in prioriteiten tussen hun behandelaar en henzelf.

Deze factoren werden genoemd als belangrijke barrières voor adequaat kunnen omgaan met hun nieuwe diagnose. Veel patiënten benadrukten het belang van het implementeren van persoonsgerichte zorg. In deze thesis beargumenteren wij dat predictiemodellen, zowel de traditionelere modellen voor ESKD bij CKD, en overlijden bij ESKD, maar ook voor de uitkomsten genoemd in **Hoofdstuk 2**, een belangrijke rol kunnen spelen in de implementatie van persoonsgerichte zorg in zorg voor patiënten met CKD.

In **Hoofdstuk 3** kijken we in detail naar de methode die in **Hoofdstuk 2** gebruikt werd: het systematisch reviewen en thematisch analyseren van kwalitatieve studies. Deze methode kan beschouwd worden als een 'kwalitatieve meta-analyse', waarbij abstractie van deze studies wordt verkregen die verder gaat dan de individuele studies. Kwalitatieve data, en de interpretatie van de auteurs, worden geëxtraheerd en dan thematisch geanalyseerd op een methode die vergelijkbaar is met kwalitatieve studies. Ons eerste doel was descriptief: tussen de eerste kwalitatieve review uit 2007, en de 2020 werden 1.695 methodologisch vergelijkbare reviews gepubliceerd. Wij beschrijven de belangrijkste karakteristieken van deze studies, en demonstreren een exponentiele groei van deze relatief nieuwe onderzoeksmethode.

Voor het beoordelen van kwalitatief onderzoek bestaan meerdere richtlijnen. De COREQ (Consolidated criteria for reporting qualitative research)[3], en de ENTREQ (Enhancing transparency in reporting the synthesis of qualitative research), respectievelijk voor interview- en focusgroepstudies, en voor kwalitatieve reviews zijn de meest frequent gebruikte. Ons tweede doel was de rapportagekwaliteit van de individuele studies (bepaald d.m.v. de COREQ), en het gebruik van de ENTREQ te onderzoeken. Een opvallende bevinding was de significante verbetering van kwalitatieve studies sinds de publicatie van

de COREQ, al blijven enkele van de items sterk achter ten opzichte van de rest.

A1.2.2 Samenvatting van deel 2

Het tweede deel van deze thesis focust op het voorspellen van een toekomstig IS over het gehele spectrum van de nierfunctie. In **Hoofdstuk 4** identificeren we middels een systematische review 77 modellen die ontwikkeld zijn voor deze uitkomst, waarvan we 15 modellen naast elkaar extern valideerden in een incidente dialysepopulatie; een zgn. *head to head* vergelijking. Deze modellen bleken slecht te functioneren in deze externe validatie: de discriminatie was matig (de c-index varieerde van 0.49 tot 0.66), en de modellen over- of onderschatten de risico's sterk. De CHA_2DS_2-VASc, een frequent gebruikt model dat o.a. door de ESC richtlijnen wordt geadviseerd, had een lage c-index (0.65 [95% CI 0.57-0.73]), en was slecht gekalibreerd.

In **Hoofdstuk 1**, de algemene inleiding, werd besproken dat het risico IS bij CKD toeneemt in combinatie met atriumfibrilleren (AF). Daarom richtte **Hoofdstuk 5** zich op patiënten met AF en, in drie gestratificeerde groepen van nierfunctie: van een normale nierfunctie tot ESKD. Gegevens van 1.372.425 inwoners van Stockholm met bekende creatininewaarden werden gebruikt om 36.004 patiënten met nieuw ontstane AF te selecteren en te categoriseren volgens KDIGO-stadia. Zes predictiemodellen (de AFI, CHADS, Modified-$CHADS_2$, CHA_2DS_2-VASc, ATRIA en GARFIELD-AF) werden extern gevalideerd, opnieuw in een head to head vergelijking. De meeste modellen vertoonden een negatief verband tussen discriminatie en afnemende nierfunctie, en bleken slecht gekalibreerd, ongeacht het KDIGO-stadium. De ideale '*prediction horizon*', de periode tussen het moment van voorspellen en de uitkomst, lag voor veel modellen op een ander tijdstip dan in de ontwikkelingsstudie. Over het geheel genomen presteerde het Modified-$CHADS_2$-model goed, met stabiele discriminatie (c-index tussen 0.73 en 0.78) en goede kalibratie, vooral in vergelijking met de CHA_2DS_2-VASc.

Head-to-head vergelijkingen van meerdere modellen in een relevante doelpopulatie is cruciaal voor implementatie. We laten daarmee in **Hoofdstukken 4 en 5** beduidend lagere c-indices zien in vergelijking met de oorspronkelijke ontwikkelingspopulatie. Dit kan deels worden toegeschreven aan andere factoren dan het model zelf, zoals variatie in de samenstelling van de ontwikkelings- en de validatiepopulatie. Echter, methodologische tekortkomingen en bias hebben waarschijnlijk ook bijgedragen aan de teleurstellende prestaties van de meeste modellen in deze hoog-risico populatie. In **Hoofdstuk 6** gaan we hier dieper op in, en bespreken we het concept van ROB in voorspellend onderzoek en introduceren we de Prediction model Risk Of Bias ASsessment Tool (PROBAST). In

dit hoofdstuk bieden we handvatten om de methodologische kwaliteit van voorspellende modellen te beoordelen, bespreken we methoden voor het beoordelen van modellen in reviews, en dragen we suggesties aan om modelontwikkeling te verbeteren. We onderstrepen de noodzaak van een dergelijke ROB tool aan de hand van een meta-review: gemiddeld was de ROB onduidelijk of hoog in alle vier domeinen, maar vooral het Analyse-domein.

A1.3 INTERPRETATIE VAN BEVINDINGEN EN IMPLICATIES VOOR ONDERZOEK EN KLINISCHE PRAKTIJK

Beide delen van deze thesis bestonden uit een toegepast gedeelte en een methodologisch gedeelte. In deze sectie bespreken we de implicaties die dit onderzoek kan hebben voor zowel de clinicus als onderzoeker.

A1.3.1 Implicaties voor de klinische praktijk.

A1.3.1.1 Risico op ischemische beroerte bij CKD

In de toegepaste gedeelten van **deel 1 en 2** hebben we het perspectief van patiënten met CKD op persoonsgerichte zorg onderzocht, evenals de voorspellende capaciteiten van predictiemodellen voor IS in CKD-zorg. Hoewel deze onderwerpen op het eerste gezicht niet direct gerelateerd lijken, beargumenteren we dat modellen nuttig zijn binnen persoonsgerichte zorg door het berekenen van risico's die specifiek zijn voor de patiënt. In **Hoofdstuk 2** hebben we lieten we zien dat het bereiken van ESKD een van de belangrijkste uitkomst was voor patiënten met CKD. Andere aspecten – of uitkomsten – zoals ervaren gezondheid, sociale en economische aspecten van leven met een ziekte, werden echter ook genoemd en waren voor veel patiënten minstens zo belangrijk.

Helaas werden deze mogelijke uitkomsten en bijbehorende risico's zelden besproken door behandelaars. Hiervoor zijn twee duidelijke barrières aan te wijzen. Ten eerste was er een duidelijke mismatch tussen de geprioriteerde uitkomsten van patiënten en hun medisch team, bijvoorbeeld door een informatieachterstand van patiënten, of een gebrek aan kennis van hun behandelaars. Patiënten pleitten voor holistische en persoonsgerichte zorg. De tweede reden is de onzekerheid van de risicoschattingen voor deze andere uitkomsten, waardoor artsen een aarzeling voelden om deze met patiënten te bespreken. Desondanks gaven patiënten duidelijk aan dat zij wel geïnformeerd willen worden. Vanuit methodologisch standpunt kunnen veel van deze aspecten van leven met CKD als uitkomsten worden beschouwd waarvoor mogelijk risico's kunnen worden berekend. Hoewel we al stappen hebben gezet, zoals het identificeren van 42 modellen voor het voorspellen van ESKD en

16 modellen voor het risico op overlijden bij ESKD in eerdere studies, blijven veel door patiënten en zorgverleners geprioriteerde uitkomsten nog ongemodelleerd. Dit benadrukt de uitdaging, maar ook het potentieel van predictieonderzoek in dit vakgebied.

In deze scriptie hebben we ons toegespitst op een specifieke uitkomst, namelijk IS – een van de door patiënten benadrukte uitkomst – en zo de brug geslagen tussen beide toegepaste delen. Modellen voor IS zijn traditioneel ontwikkeld in populaties van patiënten met AF, maar hun validiteit in de CKD-populatie is nog onduidelijk. Zoals we illustreren in **Figuur 1.2** is het risico op IS bij CKD complex en omvat klassieke risicofactoren en ziekte-specifieke factoren, naast een verhoogd risico op het ontwikkelen van AF, wat ook sterk aan IS gerelateerd is. Daarnaast is het risico op (therapie-gerelateerde) bloedingen sterk verhoogd. Hoewel richtlijnen zoals ESC en AHA/ASA deze verhoogde risico's noemen, zijn hun aanbevelingen voor patiënten met CKD gelijk aan die voor algemene patiënten met AF: de CHA_2DS_2-VASc-score wordt aanbevolen voor risicoschatting. Er bestaan echter ook andere IS-risicomodellen die minder aandacht hebben gekregen dan de conventioneel veelgebruikte CHA_2DS_2-VASc. En gezien de matige prestaties van dit model over het gehele spectrum van de nierfunctie, is er potentieel voor het identificeren van een beter presterend model voor CKD en dialysepatiënten. Dit wordt geïllustreerd in **Hoofdstukken 4 en 5**: hoewel de CHA_2DS_2-VASc het best presterende model was in onze head-to-head validatiestudie bij dialysepatiënten (**Hoofdstuk 4**), presteerde het nog steeds matig met een c-index van 0.65 (95% betrouwbaarheidsinterval 0.57-0.73). Het discriminerend vermogen was vergelijkbaar met de minder bekende modified-$CHADS_2$ (c-index van 0.62 (0.56-0.68)), dat het beste en meest stabiele presterende model was over het gehele spectrum van nierfunctie in **Hoofdstuk 5** (c-index 0.78 (0.77-0.79) bij normale eGFR, 0.73 (0.71-0.74) bij milde CKD, en 0.74 (0.69-0.79) bij gevorderde CKD zonder dialyse). De CHA_2DS_2-VASc presteerde slechter: bij een normale eGFR, milde en gevorderde CKD was de c-index respectievelijk 0.70 (0.69-0.71), 0.60 (0.58-0.62) en 0.58 (0.52-0.64). Vergeleken met de beter functionerend Modified-$CHADS_2$ verhoogt het gebruik van de CHA_2DS_2-VASc dus het risico op misclassificatie: het verkeerd inschatten van het risico op IS in deze hoog-risicopopulatie. Het risico bestaat dat hierdoor hoog-risicopatiënten onterecht een lage kans, en omgekeerd laag-risico patiënten een te hoge kans op een IS voorspeld krijgen. Als de clinicus zich laat leiden door deze predictiemodellen zal bij de onderschatte risico's onterecht antistolling worden onthouden, en bij de patiënten met een overschat risico juist wel antistolling worden voorgeschreven. Hierdoor is bij de eerste groep het risico op IS, en bij de tweede groep de kans op therapie-gerelateerde bloedingen juist hoger.

A1.3.1.2 Het wegen van risico's: ischemische beroerte en bloeding

Naast discriminatie is kalibratie, oftewel de overeenstemming tussen voorspelde en geobserveerde risico's, een cruciale kwaliteitsparameter van externe validatiestudies. Gangbare kalibratiemethoden zijn o.a. *calibraton in the large* (de gemiddelde geobserveerde en voorspelde risico's), de Hosmer-Lemeshow-test (vergelijking van voorspelde risico's met waargenomen gebeurtenissen via chi-kwadraat) en de *calibration slope* (regressie van geobserveerde resultaten op voorspelde risico's). Kalibratie kan ook worden gevisualiseerd met een kalibratiegrafiek waarin geobserveerde en voorspelde risico's worden weergegeven.

Naast het sterk verhoogde risico op IS is bij patienten met CKD het risico op (therapie-gerelateerde) bloedingen sterk verhoogd en wordt er om deze reden vaak afgezien van behandeling met antistolling. De ESC guideline adviseert de HAS-BLED te gebruiken om het bloedingsrisico in te schatten, maar ongeacht het bloedingsrisico te starten met antistolling. In situaties waarin het starten of stoppen van behandeling afhangt van voorspelde risico's, is nauwkeurige kalibratie cruciaal om te voorkomen dat de behandeling meer kwaad dan goed doet: de behandelaar dient adequaat gekalibreerde risico's op IS en op bloedingscomplicaties met elkaar te vergelijken. Helaas ligt in validatiestudies de nadruk vaak op discriminatie, en wordt informatie over kalibratie vaak niet gerapporteerd, wat resulteert in beperkte gegevens over kalibratie. De beperkte informatie die beschikbaar is, en waar **hoofdstukken 4 en 5** aan bijgedragen heeft, suggereert dat de meeste modellen, inclusief de CHA_2DS_2-VASc het risico op IS substantieel over- of onderschat. De Modified-$CHADS_2$ daarentegen laat accurate en stabiele kalibratie zien. Er zijn weinig validatiestudies van modellen die het bloedingsrisico voorspellen, maar op basis van deze data kan een vergelijkbare conclusie getrokken worden over de conventioneel gebruikte bloedingsmodellen. Een belangrijke klinische implicatie van deze thesis is de aanbeveling de Modified-$CHADS_2$ te gebruiken voor IS risicoschatting bij patiënten met CKD. Voor het accuraat inschatten van bloedingsrisico dient verder onderzoek verricht te worden.

A1.3.2 Methodologische implicaties

Dit proefschrift heeft twee methodologische secties, waarbij aan de hand van rapportage- en ROB-richtlijnen methodologische aspecten van respectievelijk kwalitatief en predictieonderzoek bespreken. In **Hoofdstuk 3** evalueren we de impact van de ENTREQ en COREQ richtlijnen op de kwaliteit van rapportage in kwalitatieve reviews. En in **Hoofdstuk 6** bespreken we de PROBAST, een ROB-richtlijn voor predictieonderzoek. Rapportage- en ROB-richtlijnen hebben veel overeenkomsten, maar ook belangrijke verschillen die in deze sectie besproken worden.

A1.3.2.1 Rapportagerichtlijnen: de COREQ

Transparant en reproduceerbaar rapporteren van onderzoeksresultaten is een essentieel onderdeel van wetenschappelijk onderzoek. Richtlijnen voor rapportage kunnen hier aan bijdragen: ze bieden onderzoekers een overzicht van onderdelen die gerapporteerd dienen te worden in hun artikel, reviewers en redacties handvatten om te controleren of alle noodzakelijke informatie in het artikel beschreven wordt, en faciliteren door het volledig en eenduidig rapporteren van data het gebruik van artikelen in meta-analyses. Daarnaast biedt het onervaren onderzoekers een kapstok waarmee veelgemaakte fouten en valkuilen vermeden kunnen worden, vaak ook door het geven van tal van voorbeelden over hoe een artikel geschreven kan worden. Een van de belangrijkste argumenten tegen het gebruik is dat richtlijnen voor rapportage beperkend kunnen zijn. De richtlijnen kunnen auteurs verplichten om specifieke opmaakvereisten te volgen, wat hun creativiteit kan beperken.

Richtlijnen voor het beoordelen van de rapportage verschillen echter fundamenteel van ROB-richtlijnen waar doorgaans methodologische kwaliteit mee onderzocht wordt. In de praktijk is deze scheidslijn minder duidelijk, en worden rapportagerichtlijnen ook gebruikt als toets voor de kwaliteit. Dit tonen wij in aan in **Hoofdstuk 3**, waarin we het gebruik van de COREQ – een richtlijn voor rapportage – door auteurs van kwalitatieve reviews exploreren. Van de 1.695 reviews gebruikten 284 (14%) de COREQ; een deel hiervan gebruikte deze richtlijn ook als beslishulpmiddel voor het includeren van artikelen. Hoewel het niet rapporteren van items een reden kan zijn om de methodologische kwaliteit van een artikel verder tegen het licht te houden, zijn er genoeg valide redenen om deze items niet te rapporteren. Zo is de COREQ specifiek ontwikkeld voor interview- en focusgroepstudies: bij het gebruik hiervan voor methodologisch andere studies zal een deel van de items niet gerapporteerd kunnen worden.

Opvallend was dat sommige items belangrijker beschouwd dan andere: van de 32 items werden slechts 18 voor ten minste 90% van de beoordelingen gebruikt. Daarnaast combineerde een substantieel deel van de reviews de COREQ met andere richtlijnen voor kwalitatief onderzoek. Dit illustreert het onopgeloste debat over het gebruik van rapportagerichtlijnen in kwalitatief onderzoek, waar ons artikel door het aggregeren van de feitelijke data over het gebruik hiervan, een belangrijke impuls aan geeft.

A1.3.2.2 ROB in predictie: de PROBAST

Er bestaat weinig empirisch bewijs over het effect van bias in predictieonderzoek. De PROBAST definieert bias in deze setting als:

> "(…) when shortcomings in the study design, conduct, or analysis lead to systematically distorted estimates of model predictive performance."

Bias resulteert, volgens deze definitie, dus in een foutieve schatting van de externe validiteit van het model. Deze hypothese is recent getoetst in een validatiestudie van de PROBAST, waarbij inderdaad een negatief effect op de discriminatie werd gezien bij validatiestudies van modellen met een hoge ROB. In deze studie worden verschillende punten van kritiek op de PROBAST geuit: de methode is te complex (de complexiteit wordt overigens erkend door de PROBAST-auteurs, die stellen dat "both subject and methodological expertise" vereist is), en de inter-beoordelaarsvariatie lijkt hoog te zijn, vermoedelijk zowel door de multi-interpretabele 'signaling questions' (SQ), als door de scoringsmethodiek.

Wij demonstreren daarnaast dat de PROBAST matig kan differentiëren tussen hoge- en zeer hoge ROB. De PROBAST bestaat uit vier domeinen (participants, predictors, outcome and analysis), en 20 SQ die met 'ja', 'waarschijnlijk ja', 'waarschijnlijk nee', 'nee' en 'geen informatie' beantwoord kunnen worden – vijf opties per SQ dus. ROB per domein is geclassificeerd als 'hoog' als één of meer SQ met 'nee' wordt beantwoord. Probabilistisch resulteert in een grotere kans op hoge ROB voor domeinen met meer SQ (respectievelijk twee, drie, zes en negen SQ voor de vier domeinen), iets wat wij inderdaad demonsteren in **Hoofdstuk 6**. In een vervolgstudie, niet in deze thesis geïncludeerd, repliceren wij deze bevindingen en tonen aan dat er waarschijnlijk sprake is van een 'ceiling effect': de PROBAST is niet goed in staat onderscheid te maken tussen modellen met een hoog en zeer hoog ROB. Wat hier aan bijdraagt is dat elke SQ voor de geaggregeerde domeinscore een gelijk gewicht heeft. Het is echter methodologisch lastig te verdedigen dat bijv. SQ 3.5 ("Was the outcome determined without knowledge of predictor information?") een vergelijkbare impact zal hebben als bijv. SQ 4.4 ("Were participants with missing data handled appropriately?").

De PROBAST is momenteel een veelgebruikte, en vermoedelijk de meest bekende methode om ROB in te schatten. In **Hoofdstuk 6**, gericht op clinici en onderzoekers met interesse in de methodologie van predictiemodellen. Hier bespreken we de afzonderlijke SQ, en gaan dieper in op enkele aspecten die niet in de PROBAST worden behandeld. Relevant voor **Hoofdstuk 5** is het effect van de predictie-horizon: idealiter voorspellen prognostische modellen een uitkomst binnen een bepaalde termijn, bijvoorbeeld één jaar voor de CHA_2DS_2-VASc of vijf voor de Modified-$CHADS_2$. Als, in de statistische analyse, deze predictie-horizon niet wordt gedefinieerd, zal de gehele follow-up periode van het cohort gebruikt worden. Afhankelijk van de voorspelde uitkomst zal het geobserveerde risico op deze uitkomst toenemen met het verlengen van deze periode. In **Hoofdstuk 5** onderzoeken

we het effect van het sequentieel verlengen van de follow-up periode op discriminatie en kalibratie. We demonstreren dat discriminatie redelijk stabiel blijft, terwijl de kalibratie door het toenemende geobserveerde risico sterk kan afwijken.

Met de stabiele prevalentie van hoge ROB (hetzij als reflectie van het eerder genoemde *ceiling-effect*, hetzij als werkelijk stabiele trends), in een steeds groeiend veld van predictieonderzoek, blijft de PROBAST momenteel de enige beschikbare tool om ROB te beoordelen. Ondanks de hierboven benadrukte tekortkomingen, en de complexiteit van het instrument, blijft de het een nuttig instrument om methodologische tekortkomingen in predictiestudies te beoordelen. Onze studies over dit onderwerp kunnen bijdragen aan een bijgewerkte versie van de PROBAST, of aan een geheel nieuw ROB-hulpmiddel, waarbij rekening wordt gehouden met de genoemde overwegingen.

A1.4 IMPLICATIES EN AANBEVELINGEN VOOR VERDER ONDERZOEK

In de bovenstaande secties hebben we de patiëntenperspectieven op CKD-zorg en predictieonderzoek, de validiteit van modellen voor IS bij patiënten met CKD, en het gebruik en rapportage- en ROB-checklists in kwalitatief en predictieonderzoek besproken. In deze laatste sectie bespreken we de implicaties van dit onderzoek, en geven we suggesties voor verder onderzoek.

A1.4.1 CKD: luister naar de patiënt

Patiënten met CKD ervaren onzekerheid over welke uitkomsten zich kunnen voordoen en wanneer deze kunnen optreden. Dit vormt de kern van **Hoofdstuk 2** en benadrukt het belang van het adresseren van deze vragen in zowel klinische zorg als onderzoek. Deze uitkomsten omvatten diverse aspecten zoals fysieke en mentale symptomen, maar ook sociale impact en economische productiviteit, waarbij opvalt dat patiënten vaak andere prioriteiten stellen dan hun zorgverleners. Patiënten met CKD voelen zich vaak onvoldoende geïnformeerd over de uitkomsten van CKD en benadrukken de behoefte aan holistische zorg die rekening houdt met hun behoeften. Het gebruik van door patiënten gerapporteerde uitkomstmaatregelen (PROMs) kan hierbij helpen en heeft veelbelovende resultaten getoond in klinieken waar ze zijn geïmplementeerd.

Patiënten benadrukken ook de behoefte aan open communicatie over prognose, ongeacht de prognostische onzekerheden. Hoewel voorspellingsmodellen in CKD-zorg voornamelijk gericht zijn op het voorspellen van ESKD of overlijden, kan het voorspellen van andere relevante uitkomsten, zoals die besproken in **Hoofdstuk 2**, de zorg meer patiëntgericht

maken. Een veelbelovende ontwikkeling is het voorspellen van toekomstige PROMs. Deze thesis illustreert het gebrek aan, maar ook het potentieel van predictieonderzoek in dit vakgebied.

A1.4.2 Kwalitatief onderzoek: heldere rapportage is essentieel

Het gebruik van kwalitatieve onderzoeksmethoden in medisch onderzoek groeit, evenals het aantal reviews van kwalitatieve studies. **Hoofdstuk 3** onderzocht het gebruik van rapportagerichtlijnen zoals COREQ en ENTREQ in deze reviews. Na de introductie van COREQ verbeterde de kwaliteit van de rapportage. Deze bevindingen hebben implicaties op drie niveaus: kwalitatief onderzoek in het algemeen, reviews van kwalitatief onderzoek en originele studies. Ten eerste is het gebruik van rapportagerichtlijnen in kwalitatief onderzoek niet onbetwist. Door feitelijke gegevens te verstrekken over het gebruik van deze checklists en het effect van de publicatie van de COREQ op de kwaliteit van de rapportage, geven onze bevindingen een impuls aan deze discussie. Ten tweede: onze review geeft inzicht in het gebruik van de COREQ als ROB tool in reviews, waar het nadrukkelijk niet voor ontwikkeld is. Ten slotte zouden auteurs van een kwalitatieve studie hun artikel extra kunnen controleren aan de hand van de items van de COREQ die slecht scoorden in onze review, om zo het risico op vergelijkbare slechte rapportage te minimaliseren.

A1.4.3 Antistolling in CKD: weeg het risico op IS en bloeding

Vanuit klinisch en methodologisch oogpunt is verder onderzoek essentieel om het risico op IS en bloedingen bij CKD- en dialysepatiënten nauwkeurig in te schatten. Antistolling wordt minder vaak voorgeschreven aan deze kwetsbare populatie vanwege de angst voor therapie-gerelateerde bloedingen. Een gepersonaliseerde aanpak waarbij het risico op zowel IS als bloedingen wordt berekend en afgewogen, vervolgens wordt besproken met de patiënt, voordat antistolling wordt gestart, is essentieel. Om dit te bereiken, kan net als in deze thesis een overzicht worden gemaakt van alle bestaande bloedingsrisicomodellen, gevolgd door externe validatie van deze scores in relevante CKD- en dialysecohorten. Hierbij dient speciale aandacht te worden besteed aan kalibratie, wat vergelijking van IS- en bloedingsrisico mogelijk maakt, het risico op misclassificatie vermindert en dus incorrecte behandelbeslissingen voorkomt.

A1.4.4 Predictieonderzoek: focus op methodologie

Predictiemodellen met een hoog ROB presteren slecht in externe validatiestudies. De definitie van bias in predictieonderzoek is complex, en hoewel de PROBAST helpt in het identificeren van bias, en er intussen veel methodologische publicaties over dit onderwerp zijn verschenen, blijft de methodologische kwaliteit van veel nieuwe modellen matig. Dit kan het gevolg zijn van de kwaliteit van het meetinstrument – de PROBAST – zelf: we demonstreerden een *ceiling effect*, en bespraken verschillende punten waarin de PROBAST verbeterd kan worden.

A1.4.5 Alternatieve strategieën voor modelontwikkeling

Er worden talloze modellen ontwikkeld. Veel van deze modellen worden echter niet extern gevalideerd en zullen waarschijnlijk nooit worden gebruikt in de klinische praktijk. In plaats van nieuwe modellen te ontwikkelen, kan het valideren van bestaande en veelgebruikte modellen in verschillende doelpopulaties een haalbaar alternatief zijn. We hebben aangetoond dat *head-to-head validations,* het valideren van meerdere modellen in dezelfde populaties, waardevolle informatie oplevert waarmee het best presterende model in deze populatie geïdentificeerd kan worden. Indien nodig kan dit model vervolgens worden geherkalibreerd om de prestaties verder te verbeteren. Ten slotte hebben we het effect van de *prediction horizon* op discriminatie en kalibratie onderzocht. We toonden aan dat de optimale kalibratie vaak op een andere *prediction horizon* ligt dan in de originele studies geadviseerd wordt. Het samenvoegen van meerdere modellen kan een interessante, maar nog niet geëxploreerde optie zijn om een uitkomst nauwkeurig te voorspellen op verschillende tijdsintervallen.

A1.5 CONCLUSIE

Concluderend onderzoekt deze thesis de perspectieven van patiënten met CKD op prognose als onderdeel van persoonsgerichte zorg en illustreert de rol die predictiemodellen hierin kan vervullen. Daarnaast geeft het een methodologische basis voor zowel de kwalitatieve als prognostische onderzoeksmethodologie. We hebben een kwalitatieve basis gecreëerd die zowel gebruikt kan worden in klinische zorg als in onderzoek, door een overzicht te creëren van geprioriteerde uitkomsten van CKD, wat als basis gebruikt kan worden voor de ontwikkeling modellen die voor patiënten relevant zijn. Als illustratie hiervan onderzochten we de validiteit van modellen voor IS in CKD, en demonstreerden de matige externe validiteit hiervan. De Modified-$CHADS_2$, een eenvoudige risicoscore, liet over het gehele spectrum van de nierfunctie goede discriminatie en kalibratie zien, en zou een goed alternatief zijn voor de CHA_2DS_2-VASc voor risicostratificatie in deze hoog-risicopopulatie.

APPENDIX 2: ACKNOWLEDGEMENTS

Dankwoord

Dit proefschrift had niet tot stand kunnen komen zonder de begeleiding en ondersteuning van een aantal personen, die ik hier dan ook graag bedank.

Allereerst wil ik mijn promotor, Prof. Dr. Friedo Dekker en copromotor Dr. Merel van Diepen bedanken. Friedo, jouw constante aanmoediging, kritische feedback en de tijd die je in mijn begeleiding hebt geïnvesteerd zijn van grote waarde geweest voor de totstandkoming van deze thesis, maar ook voor mijn eigen academisch ontwikkeling. Ik ben dankbaar voor de vrijheid die ik van jou gekregen heb om mijn eigen weg te zoeken.

Merel, jouw begeleiding gedurende deze jaren, zowel inhoudelijk als op persoonlijk vlak, was zeer waardevol. Jouw kennis, scherpe inzichten, geduldige begeleiding en humor, hielpen mij de methodologie eigen te maken, nieuwe projecten te verkennen, en hebben mijn passie voor onderzoek versterkt.

Verder wil ik de onderzoekers en artsen bedanken waarmee ik heb mogen samenwerken, in het bijzonder Dr. Yvette Meuleman, Dr. Gurbey Ocak, Prof. Dr. Willem Jan Bos, Prof. Dr. Rolf Groenwold, Dr. Maarten van Smeden, Prof. Dr. Ewout Steyerberg, Prof. Dr. Joris Rotmans, en Prof. Dr. Kitty Jager. I would like to thank Dr. Marco Trevisan, Dr. Karolina Szummer and Prof. Dr. Carmine Zoccali for their collaboration.

I'd like to express my gratitude to Prof. Dr. Rachael Morton for her time to meet me in Budapest and later in Oxford where we discussed methodological aspects of qualitative research. Additionally, I extend my thanks to Prof. Dr. Juan Jesus Carrero for granting me the opportunity to work with the SCREAM dataset, and for our productive discussions during the 2019 ERA congress.

Daarnaast wil ik de collega's van de afdeling Klinische Epidemiologie bedanken. Tamara en Yvonne, jullie waren altijd meedenkend en behulpzaam. De stafleden dank ik voor het uitstekende onderwijs, en de leden van Nefro Journal Club voor de wekelijks terugkerende inspirerende start van de week. Esmee, dank voor je vele uren samenwerking en evenveel koffiemomenten en goede gesprekken. Chava, dank dat je me hebt ingewerkt op zowel de afdeling als de materie. Bas, dank voor het vlottrekken van de LOWESS-curves en de gesprekken hoe onze *counterfactual* alter-ego's als niet-promovendus zouden functioneren. Monique, dank dat ik je 'epi-buddy' mocht zijn; een functie voor het leven. Verder bedank ik Edouard, Myrthe, Sebastiaan, Kevin en Cynthia voor de prettige samenwerking.

Ook dank ik Prof. Dr. Hans de Fijter, Prof. Dr. Natasha Appelman en drs. Eveline de Lange voor het faciliteren van dit promotietraject tijdens de opleiding tot internist.

Daarnaast wil ik de studenten die ik heb mogen begeleiden bedanken: Vera van der Endt, Jet Milders, Danielle Derksen en Liselotte Langenhuijsen. Jullie enthousiasme was stimulerend en productief.

Jan Schoones: vanaf dag één, intussen ruim 13 jaar geleden, een rode draad. Dank voor de toegang tot de literatuur, de inspiratie voor het verrichten van gedegen onderzoek en natuurlijk het klaverjassen en het abdijbier.

Gido, onze wandelgesprekken waren waardevol en soms ook helend. Het hardlopen wat minder, maar dat pakken we wel weer op.

Hans en Robin, het begon allemaal met een meta-analyse in 2013. Nu, 11 jaar later, hebben we vergelijkbare, en soms gelijktijdige routes naar verschillende eindpunten gelopen.

Saakje en Bas, dank dat jullie mijn paranimfen willen zijn op deze bijzondere dag.

Hans en Anneleen, dank voor jullie warmte en betrokkenheid. Heit en mamma, dank voor de liefdevolle basis die jullie legden.

Joris, Lotte, Thomas en Olaf: zonder jullie was ‘heitie’s boek’ misschien een hoofdstuk rijker, maar mijn leven vier heerlijke kinderen armer. Ik geniet elke dag van jullie. “Wanneer is je boek nou klaar?” – nou, hier is het dan. En Femke, zo graag had ik jou leren kennen.

Lieve Sabine, ik houd van jou.

APPENDIX 3: CURRICULUM VITAE

Ype de Jong is geboren in Amersfoort op 2 april 1990. Hij behaalde zijn diploma's VMBO, HAVO en VWO op verschillende scholen in Groningen, waarna hij in 2009 verhuisde naar Leiden om geneeskunde te studeren in het Leids Universitair Medisch Centrum (LUMC). Tijdens zijn studie publiceerde hij een meta-analyse over de ideale lichaamshouding voor mannen om te urineren. Een project dat begon als een studentikoze grap, maar uiteindelijk leidde tot de publicatie van twee artikelen en het ontvangen van twee scriptieprijzen. Maar bovenal resulteerde het in waardevolle onderzoekservaring en het aanwakkeren van interesse in wetenschappelijk onderzoek en onderzoeksmethodologie. Aan het einde van zijn master volgde hij als eerste student in Leiden het dedicated schakeljaar interne geneeskunde, en een maand na zijn afstuderen (*Cum Laude)* begon hij in oktober 2015 met de opleiding tot internist.

Na een klinische periode van twee jaar in het HMC in Den Haag begon hij met een driejarig promotietraject op de afdeling klinische epidemiologie onder begeleiding van Dr. Merel van Diepen en Prof. Dr. Friedo Dekker. Dit promotietraject bestond uit kwalitatief onderzoek, validatiestudies van predictiemodellen voor ischemische beroerte bij mensen met een verminderde nierfunctie, en methodologisch onderzoek naar publicatierichtlijnen en risk of bias tools binnen kwalitatief- en predictie onderzoek. Tijdens het promotietraject volgde hij verschillende cursussen in het kader van de opleiding tot Epidemioloog B, gaf hij onderwijs aan (bio)medische studenten en promovendi, en presenteerde hij zijn resultaten op verschillende internationale congressen. Daarnaast behaalde hij zijn basiskwalificatie onderwijs (BKO), en begeleide hij studentprojecten met meerdere publicaties als resultaat. Sinds 2023 is hij in opleiding tot internist vasculair geneeskundige op de afdeling trombose en hemostase in het LUMC, en sinds 2024 volgt hij een tweede differentiatie tot klinisch farmacoloog. Daarnaast is hij lid van de Wetenschappelijke Adviesraad van het Geneesmiddelen-Bulletin (Ge-Bu). Als bestuurslid van de Nederlandse Vereniging van Internisten Vasculaire Geneeskunde (NVIVG) is hij verantwoordelijk voor het organiseren van nascholings- en congresdagen.

Ype is getrouwd met Sabine en samen met Joris (2018), Lotte (2018), Thomas (2020) en Olaf (2022) woont hij in Roelofarendsveen.

APPENDIX 4: COAUTHORS

This appendix lists the coauthors and their affiliation(s). The list is sorted alphabetically.

Blankestijn PJ	Chapter 4	Department of Nephrology and Hypertension, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands.
Bos WJW	Chapter 4	Department of Internal Medicine, Leiden University Medical Center, Leiden, the Netherlands; Department of Internal Medicine, Sint-Antonius Hospital, Nieuwegein, the Netherlands.
Carrero JJ	Chapter 5	Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. Department of Clinical Sciences, Karolinska Institutet, Danderyd Hospital, Stockholm, Sweden.
Dekker FW	Chapter 2,3,4,5,6	Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden, The Netherlands.
Fu EL	Chapter 5	Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, USA. Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden, The Netherlands.
Jager KJ	Chapter 6	ERA Registry, Amsterdam UMC location University of Amsterdam, Department of Medical Informatics, Amsterdam, The Netherlands. Amsterdam Public Health, Quality of Care and Ageing & Later Life, Amsterdam, The Netherlands.
Meuleman Y	Chapter 2,3	Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden, The Netherlands.
Milders J	Chapter 2,3	Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden, The Netherlands.
Morton RL	Chapter 2,3	Health Economics and Health Technology Assessment, NHMRC Clinical Trials Centre, Faculty of Medicine and Health, University of Sydney, Sydney, NSW, Australia.
Ocak G	Chapter 4,5	Department of Internal Medicine, Sint Antonius Hospital, Nieuwegein, The Netherlands.
Ramspek CL	Chapter 4,6	Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden, The Netherlands.
Rookmaaker MB	Chapter 4	Department of Nephrology and Hypertension, University Medical Center Utrecht, Utrecht, The Netherlands.

Szummer K	Chapter 5	Department of Cardiology, Heart and Vascular Center, Karolinska University Hospital, Stockholm, Sweden. Department of Medicine Huddinge, Karolinska Institutet, Stockholm, Sweden.
Trevisan M	Chapter 5	Department of Medical Epidemiology and Biostatistics, Campus Solna, Karolinska Institutet, Stockholm, Sweden.
van der Endt VHW	Chapter 4	Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden, The Netherlands.
van der Willik EM	Chapter 2,3	Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden, The Netherlands. Department of Epidemiology and Data Science, Amsterdam UMC location Vrije Universiteit, P.O. box 7057, Amsterdam, 1007 MB, the Netherlands.
van Diepen M	Chapter 2,3,4,5,6	Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden, The Netherlands.
Verhaar MC	Chapter 4	Department of Nephrology and Hypertension, University Medical Center Utrecht, Utrecht, the Netherlands.
Vernooij RWM	Chapter 4	Department of Nephrology and Hypertension, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands.
Voorend CGN	Chapter 3	Department of Internal Medicine (Nephrology), Leiden University Medical Center, Leiden, The Netherlands.
Zoccali C	Chapter 6	IPNET, c/o Nefrologia delGrande Ospedale Metropolitano, Reggio Calabria, Italy.

APPENDIX 5: LIST OF PUBLICATIONS

5.1 Published articles included in this thesis

1. **de Jong Y**, van der Willik EM, Milders J, Meuleman Y, Morton RL, Dekker FW, van Diepen M. Person centred care provision and care planning in chronic kidney disease: which outcomes matter? A systematic review and thematic synthesis of qualitative studies. BMC Nephrol. 2021 Sep 13;22(1):309. doi: 10.1186/s12882-021-02489-6. PMID: 34517825; PMCID: PMC8438879.
2. **de Jong Y**, van der Willik EM, Milders J, Voorend CGN, Morton RL, Dekker FW, Meuleman Y, van Diepen M. A meta-review demonstrates improved reporting quality of qualitative reviews following the publication of COREQ- and ENTREQ-checklists, regardless of modest uptake. BMC Med Res Methodol. 2021 Sep 12;21(1):184. doi: 10.1186/s12874-021-01363-1. PMID: 34511068; PMCID: PMC8436506.
3. **de Jong Y**, Ramspek CL, van der Endt VHW, Rookmaaker MB, Blankestijn PJ, Vernooij RWM, Verhaar MC, Bos WJW, Dekker FW, Ocak G, van Diepen M. A systematic review and external validation of stroke prediction models demonstrates poor performance in dialysis patients. J Clin Epidemiol. 2020 Jul;123:69-79. doi: 10.1016/j.jclinepi.2020.03.015. Epub 2020 Mar 30. PMID: 32240769.
4. **de Jong Y**, Ramspek CL, Zoccali C, Jager KJ, Dekker FW, van Diepen M. Appraising prediction research: a guide and meta-review on bias and applicability assessment using the Prediction model Risk Of Bias ASsessment Tool (PROBAST). Nephrology (Carlton). 2021 Dec;26(12):939-947. doi: 10.1111/nep.13913. Epub 2021 Jul 8. PMID: 34138495.
5. **de Jong Y**, Fu EL, van Diepen M, Trevisan M, Szummer K, Dekker FW, Carrero JJ, Ocak G. Validation of risk scores for ischaemic stroke in atrial fibrillation across the spectrum of kidney function. Eur Heart J. 2021 Apr 14;42(15):1476-1485. doi: 10.1093/eurheartj/ehab059. PMID: 33769473; PMCID: PMC8046502.

5.2 Other publications

6. **de Jong Y**, Pinckaers JHFM, Ten Brinck, RM, Lycklama à Nijeholt AAB. Invloed van mictiehouding op urodynamische parameters bij mannen: een literatuuronderzoek. Tijdschrift voor Urologie 4, 36–42 (2014). https://doi.org/10.1007/s13629-014-0008-5
7. **de Jong Y**, Pinckaers JH, ten Brinck RM, Lycklama à Nijeholt AA, Dekkers OM. Urinating standing versus sitting: position is of influence in men with prostate enlargement. A systematic review and meta-analysis. PLoS One. 2014 Jul 22;9(7):e101320. doi: 10.1371/journal.pone.0101320. PMID: 25051345; PMCID: PMC4106761.

8. **de Jong Y**, Haverkort ME, van Sorge AJ, Jansen C, Luyten GP, Joosse M. A translucent vascularised iris granuloma in a patient with secondary syphilis. Lancet Infect Dis. 2017 Mar;17(3):255-256. doi: 10.1016/S1473-3099(17)30071-3. Epub 2017 Feb 23. PMID: 28244379.
9. **de Jong Y**, de Kok BM, Thang HD, Hofstee HMA. Fatal sudden paralysis of the lower extremities. Neth J Med. 2017 Dec;75(10):458. PMID: 29256417.
10. **de Jong Y**, van Loenhout RB, Swank DJ, Jansen CL, Sorgdrager BJ. Polymicrobial bacterial pericarditis and cardiac tamponade caused by pericardial penetration of an adjustable gastric band. BMJ Case Rep. 2018 Feb 16;2018:bcr2017221589. doi: 10.1136/bcr-2017-221589. PMID: 29453210; PMCID: PMC5836605.
11. **de Jong Y**, Groeneveld, JHM, Vlasveld J. Trombotische microagniopathie. [Thrombotic microangiopathy]. Focus vasculair Mar 2018, edition 1, pages 30-41.
12. **de Jong Y**, Jansen CL, Muller AE, Haverkort ME. Een hoestende vrouw met een zieke parkiet.[A coughing female with a sick parakeet] The Journal of infectious diseases. 2018 April; 13(2):58-61.
13. Bakker OGM, de Kok BM, **de Jong Y**. Een brakende, oude man met buikpijn en koorts [A vomiting, old man with abdominal pain and fever]. Ned Tijdschr Geneeskd. 2018 Jun 21;162:D2412. Dutch. PMID: 30040266.
14. **de Jong Y**, De Rooy E, De Kok B, Thang O, Hofstee H. Acute paralyse van beide benen met fatale uitkomst [Acute fatal paralysis of both legs]. Focus vasculair, volume 3, issue 3, (2018)
15. de Rooij E, Verheul R, de Vreede M, **de Jong Y**. Cytomegalovirus infection with pulmonary embolism, splenic vein thrombosis and monoclonal gammopathy of undetermined significance: a case and systematic review. BMJ Case Rep. 2019 Mar 4;12(3):e226448. doi: 10.1136/bcr-2018-226448. PMID: 30837233; PMCID: PMC6424189.
16. Ramspek CL, **de Jong Y**, Dekker FW, van Diepen M. Towards the best kidney failure prediction tool: a systematic review and selection aid. Nephrol Dial Transplant. 2020 Sep 1;35(9):1527-1538. doi: 10.1093/ndt/gfz018. PMID: 30830157; PMCID: PMC7473808.
17. **de Jong Y**, Bilsen M, Grauss RW. Een man met koortsafhankelijke ecg-afwijkingen [A man with fever-dependent ECG abnormalities]. Ned Tijdschr Geneeskd. 2019 Feb 27;163:D3325. Dutch. PMID: 30816649.
18. De Rooij E, Verheul R, de Vreede M, **de Jong Y**. Veneuze trombo-embolieën verklaard door cytomegalovirus en monoklonale gammopathie van onbekende betekenis. [Venous thrombo-embolism caused by cytomegalovirus infection and monoclonal gammopathy of undetermined significance] Focus Vasculair (1); p45-51 (2020)

19. Fu EL, Janse RJ, **de Jong Y**, van der Endt VHW, Milders J, van der Willik EM, de Rooij ENM, Dekkers OM, Rotmans JI, van Diepen M. Acute kidney injury and kidney replacement therapy in COVID-19: a systematic review and meta-analysis. Clin Kidney J. 2020 Sep 2;13(4):550-563. doi: 10.1093/ckj/sfaa160. PMID: 32897278; PMCID: PMC7467593.

20. Touw CE, **de Jong Y**, van Hylckama Vlieg A. The influence of corn trypsin inhibitor on the contribution of coagulation determinants to the Technoclone Thrombin Generation Assay (TGA) and the Calibrated Automated Thrombogram (CAT). PLoS One. 2022 Feb 25;17(2):e0263960. doi: 10.1371/journal.pone.0263960. PMID: 35213588; PMCID: PMC8880747.

21. van der Endt VHW, Milders J, Penning de Vries BBL, Trines SA, Groenwold RHH, Dekkers OM, Trevisan M, Carrero JJ, van Diepen M, Dekker FW, **de Jong Y**. Comprehensive comparison of stroke risk score performance: a systematic review and meta-analysis among 6 267 728 patients with atrial fibrillation. Europace. 2022 Nov 22;24(11):1739-1753. doi: 10.1093/europace/euac096. PMID: 35894866; PMCID: PMC9681133.

22. Trevisan M, Hjemdahl P, Clase CM, **de Jong Y**, Evans M, Bellocco R, Fu EL, Carrero JJ. Cardiorenal Outcomes Among Patients With Atrial Fibrillation Treated With Oral Anticoagulants. Am J Kidney Dis. 2022 Oct 5:S0272-6386(22)00922-2. doi: 10.1053/j.ajkd.2022.07.017. Epub ahead of print. PMID: 36208798.

23. Van Paassen J, Swets MC, Groeneveld GH, de Vries JJC, **de Jong Y**, Ten Brink RM, Hiemstra PS, Zwaginga JJ, de Jonge E, Arbous MS. Viruses in the Respiratory Tract in Elective Cardiac Surgery Patients. Austin J Pulm Respir Med.2023; (10)1: 1094.

24. Langenhuijsen LFS, Janse RJ, Venema E, Kent DM, van Diepen M, Dekker FW, Steyerberg EW, **de Jong Y**. Systematic metareview of prediction studies demonstrates stable trends in bias and low PROBAST inter-rater agreement. J Clin Epidemiol. 2023 May 2;159:159-173. doi: 10.1016/j.jclinepi.2023.04.012. Epub ahead of print. PMID: 37142166.

25. van der Horst SFB, **de Jong Y**, van Rein N, Jukema JW, Palmen M, Tops LF, Klok FA, den Exter PL. Performance of Risk Scores in Predicting Major Bleeding in Left Ventricular Assist Device (LVAD) Recipients: a Comparative External Validation. *Accepted in Research and Practice in Thrombosis and Haemostasis (RPTH).*

5.3 Articles in progress

26. Van Boekel AM, van der Meijden SL, Arbous MS, Nelissen RGHH, Veldkamp KE, Nieswaag EB, Jochems KFT, Holtz J, van IJzinga Veenstra A, Reijman J, **de Jong Y**, van Goor H, Wiewel MA, Schoones JW, Geerts BF, de Boer MGJ. Systematic evaluation of machine learning models for postoperative surgical site infection prediction. *Under review PLOS ONE.*
27. Christiaans LCHH, **de Jong Y**, Mangione J, van Schie M. An Alcoholic With Reduced Consciousness and Bilaterally Rigidity Of Limbs. *Under review.*
28. Langehuijsen LSF, Derksen D, Milders J, van Diepen M, Dekker FW, Den Exter PL, Van der Horst SFB, Rotmans JI, **de Jong Y**. Meta-analysis demonstrates poor performance of stroke and bleeding prediction models in chronic kidney disease patients with atrial fibrillation. *Under review JCE.*
29. Van der Horst SFB, Chu G, Seelig J, Trinks-Roerdink EM, Voorhout L, de Vries TAC, **de Jong Y**, Klok FA, Hemels MEW, Rutten FH, Geersing GJ, Huisman MV. External Validation of the AF-BLEED Score in Predicting Major Bleeding in Patients with New-Onset Atrial Fibrillation. Under review.
30. **De Jong Y**, den Exter P. Met de rug tegen de muur: systemische trombolyse bij ernstige Splanchnicustrombose. Focus vasculair (accepted)

APPENDIX 6: EPIDEMIOLOGY PORTFOLIO

Teaching activities	Hours	Date
• Clinical appraised topic (CAT), entitled 'Ribavirin treatment in immunocompetent adults with lower respiratory tract infection due to Respiratory syncytial virus' by V.H.W. van der Endt. (Year 2, BSc Medicine)	36	01-2018
• Academische en Wetenschappelijke Vorming (Year 1, BSc Medicine)	24	03-2018
• Secondary evaluator of bachelor thesis, entitled "The quiet hospital; Earplugs to improve sleep quality in hospitalized patients" by C. van der Worp (Year 3, BSc Medicine)	2	04-2018
• Praktische OnderzoeksVaardigheden (Year 3, BSc Medicine)	4	10-2018
• Academische en Wetenschappelijke Vorming (Year 1, BSc Medicine)	24	02-2019
• Academische en Wetenschappelijke Vorming, extra werkgroep (Year 1, BSc Medicine)	4	03-2019
• Education, short SPSS course (Frontiers of Science course, FOS)	4	06-2019
• Clinical research in practice (MSc biomedical sciences)	4	09-2019
• Clinical appraised topic (CAT), entitled: "Rituximab versus cyclophosphamide for remission induction in severe ANCA-associated vasculitis" by M. Wetzels. (Year 3, BSc Medicine)	48	01-2020
• Academische en Wetenschappelijke Vorming (Year 3, BSc Medicine)	4	01-2020
• Clinical appraised topic (CAT) and bachelor thesis, entitled: 'Ischaemic stroke risk in incident dialysis patients" by J. Milders. (Year 3, BSc Medicine)	48	01-2021
• Clinical appraised topic (CAT) and bachelor thesis, entitled: 'De De CHA2DS2-VASc en chronische nierschade, een discutabele combinatie" by M. Halsema. (Year 3, BSc Medicine)	48	01-2022
• Clinical teaching activities (Year 5 MSc Medicine), as part of the BKO (basiskwalificatie onderwijs)	100	2022-2023

(Congress) presentations	Hours	Date
• Oral presentation "Nefrologie voor dummies" at dept. Clinical Epidemiology LUMC	6	08-2018
• (Invited) oral presentation, "Premature closure" at NVIVG congress Zeist.	12	09-2018
• Oral presentation: "Birth order and Delinquency: Evidence from Denmark and Florida"	6	10-2018
• Oral presentation: "Prediction models in chronic kidney disease: what do patients want to know? A systematic review and thematic synthesis of qualitative studies."	6	03-2019
• Oral presentation: "Performance of ischemic stroke risk models in incident dialysis patients: a systematic review and independent external validation study"	6	06-2019
• Poster presentation ERA EDTA Budapest 2019 "Performance of ischemic stroke risk models in incident dialysis patients: a systematic review and independent external validation study"	8	06-2019
• Oral presentation, "Measurement error in qualitative research: is it actually an issue? "	6	01-2020
• Oral presentation ERA EDTA Congress Milano 2020: "External validation of ischemic stroke risk prediction models in atrial fibrillation patients with chronic kidney disease: the Stockholm creatinine measurements (SCREAM) project"	12	06-2020

• Oral presentation: "Conducting, appraising & synthesizing qualitative research"	6	07-2020
• Oral presentation: "How to talk with conspiracy theorists?"	6	11-2020
• Oral presentation: "Risk of bias in prediction research, assessment using the PROBAST"	6	12-2020
• (Invited) oral presentation for the CVGK (Cardiovasculaire Geneeskunde), entitled: Validatie van risicoscores for ischemisch CVA in AF patiënten met CKD	12	06-2021
• (Invited) oral presentation for the PACE (Physicians' Academy for Cardiovascular Education), entitled "Validation of risk scores for ischaemic stroke in AF patients across the spectrum of kidney function"	12	06-2021
• Oral presentation for internists and nephrologists at the LUMC, entitled 'voorspellen risico iCVA bij HD patiënten"	6	06-2021
• Invited oral presentation at Nierstichting	12	10-2021
• Oral presentation for internists at the LUMC on diagnostic bias processess	6	11-2021
• Oral presentation on IS prediction within AF with or without CKD for vascular internists (LUMC)	6	10-2022
• Oral presentation on the use of post-mortem CT as opposed to autopsy (LUMC)	6	11-2023
• Oral presentation on the optimal timing of antihypertensive treatment following ischemic stroke (LUMC)	6	11-2023
• Oral presentation on the benefit of GLP1-ra in non-diabetic obese patients (LUMC)	6	11-2023
• Oral presentation on the treatment of Klippel-Trenaunay with mTOR inhibition (LUMC)	6	11-2023
• Oral presentation on the ANA-test (LUMC)	6	11-2023
• Oral presentation on thrombolytic therapy in splanchnic vein thrombosis at the NVIVG congress in Zeist	6	03-2024

Student supervision	**Hours**	**Date**
• Begeleiden V. van der Endt, keuzeonderzoek; 8 weken totaal	16	05-2018
• Begeleiden J. Milders, keuzeonderzoek; (totaal 100 weken)	200	2019
• Begeleiden V. van der Endt, keuzeonderzoek; totaal 52 weken	104	2020
• Scientific internship (wetenschapsstage), entitled: "The predictive performance of stroke and bleeding risk scores in chronic kidney disease and dialysis patients with atrial fibrillation: a systematic review and meta-analysis" by D. Derksen	60	2022
• Begeleiden L. Langehuijsen, keuzeonderzoek; totaal 2 jaar	208	2022
• Begeleiden J. van Beelen, keuzeonderzoek, totaal 6 maanden	52	2023

Courses followed / attended	**Date**
• Masterclass klinische epidemiologie te Noordwijk	2012
• Rothman-cursus klinische epidemiologie	2015
• Klinische Farmacologie	2016
• Boerhave cursus klinische communicatie	2016
• Rotterdamse Elektrolyt en zuur-basen stoornissen cursus	2017
• Water en Zout	2017
• Epidemiology, an Introduction	2018

- Basic Methods and Reasoning in Biostatistics 2018
- Survival Analysis 2018
- Statistical Aspects of Clinical Trials 2019
- Evidence Based Medicine 2019
- Regression Analysis 2019
- Onderzoeksopzet en Analyse 2019
- Gevorderde Epidemiologische Methoden 2019
- Prediction modelling and Intervention research 2019
- Klinische Epidemiologie op Schiermonnikoog 2019
- CEPD Courses Budapest 2019
- Causal inference 2020
- Capita Selecta 2020
- Meta-analyse 2020
- Klinische Genetica 2021
- eBROK 2023
- Immuniteit en Infectie 2023
- Nascholing 'CVRM Plus' 2023
- Update Hemostase 2023
- Vasculitis cursus 2023
- Jaarupdate vasculaire geneeskunde 2023
- Nascholing 'diep veneuze trombose op zeldzame(re) plekken' 2024
- Nascholing:'controverses in de vasculaire geneeskunde' 2024
- Diabetesoverleg 2024
- Diabetescursus 2024

Journal clubs / scientific discussion	**Hours**	**Date**
• Journal club, ca. 30x followed on Epi-department, Nephro-journal club (2018)	60	2018
• Journal club, ca. 40x followed on Epi-department, Nephro-journal club (2019)	80	2019
• Journal club, ca. 40x followed on Epi-department, Nephro-journal club (2020)	80	2020
• Journal club, ca. 10x followed on Epi-department, Nephro-journal club (2022)	20	2022

Awards and prizes	**Date**
• Dick Held scriptieprijs	2015
• LAG scriptieprijs	2015
• Best abstract of the congress ERA EDTA Milano	2020

Other activities	**Date**
• Researcher BCG-PRIME study	2018
• Co-chair PhD committee Dept. clinical epidemiology	2020
• Basiskwalificatie onderwijs (BKO)	2023
• Bestuurslid Nederlandse Vereniging van Internisten Vasculaire Geneeskunde (NVIVG)	
• Lid Wetenschappelijke Adviesraad (WAR) bij het Geneesmiddelen-Bulletin (Ge-Bu)	Current

A