Chronic kidney disease

Thio, C. H. L.

DOI:
10.33612/diss.133648108

CHRONIC KIDNEY DISEASE

Insights from social and genetic epidemiology

CHRIS THIO

Chris Thio

Chronic kidney disease

Insights from social and genetic epidemiology

PhD dissertation University of Groningen, the Netherlands

The publication of this thesis was financially supported by University of Groningen, University Medical Center Groningen, Science in Healthy Ageing and healthcaRE (SHARE) Research institute, and Graduate School of Medical Sciences (GSMS).

Cover design and illustrations: Senne Trip • www.sennetrip.com
Layout: Igor Wever
Print: Ridderprint | www.ridderprint.nl

ISBN: 978-94-034-2609-9 (print version)
ISBN: 978-94-034-2608-2 (electronic version)

rijksuniversiteit groningen

Chronic kidney disease

Insights from social and genetic epidemiology

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. C. Wijmenga
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

woensdag 7 oktober 2020 om 12.45 uur

door

Christian Han Liong Thio

geboren op 28 november 1983
te Hoogeveen

Promotores

Prof. dr. H. Snieder
Prof. dr. R.T. Gansevoort
Prof. dr. U. Bültmann

Beoordelingscommissie

Prof. dr. A. Koettgen
Prof. dr. P.D. Koellinger
Prof. dr. V.V.A.M. Knoers

Paranimfen

Peter van der Most
Eliza Fraszczyk
Anna Neustaeter

CONTENTS

CHAPTER 1.	General introduction and thesis outline	9
PART I		
CHAPTER 2.	Educational level and risk of chronic kidney disease: Longitudinal data from the PREVEND study	25
CHAPTER 3.	Heart rate variability and its relation to chronic kidney disease: Results from the PREVEND Study	45
PART II		
CHAPTER 4.	Familial aggregation of chronic kidney disease and heritability of renal biomarkers in the general population: The Lifelines Cohort Study	67
CHAPTER 5.	Evaluation of a genetic risk score based on creatinine-estimated glomerular filtration rate and its association with kidney outcomes	99
CHAPTER 6.	Genome-wide association scan of serum urea in European populations identifies two novel loci	117
PART III		
CHAPTER 7.	The effects of genetic factors, educational attainment, and their interaction on kidney function outcomes	139
CHAPTER 8.	Investigating causal effects of educational attainment on kidney outcomes: a Mendelian randomization study	163
CHAPTER 9.	General discussion	191
CHAPTER 10.	Nederlandse samenvatting	223
	Acknowledgements/dankwoord	235
	About the author	241
	List of publications	243
	Previous SHARE dissertations	247

General introduction and thesis outline

CHAPTER 1

Introduction

For centuries, it has been known that health disparities exist across socioeconomic groups[1]. Higher rates of disease and shorter lifespans are observed among those with lower socioeconomic status. Despite attempts to systematically reduce these disparities, they persist to this day[2,3]. These disparities are also observed for kidney disease. Those with lower education, lower income, lower occupational level, and from deprived communities, are observed to be at higher risk of chronic kidney disease (CKD)[4-6]. The mechanisms that link low socioeconomic status to CKD are not fully understood. This thesis is an effort to increase our understanding of socioeconomic disparities in kidney disease. In particular, I seek to apply modern concepts from genetic epidemiology to answer social epidemiological questions in the context of CKD. In this first chapter, I discuss background and core concepts, identify knowledge gaps, and describe aims and hypotheses. Finally, I provide an outline of this thesis.

Epidemiology of chronic kidney disease

CKD is a heterogeneous group of disorders marked by progressive loss of kidney function and/or signs of kidney damage. Currently, the international guideline group *Kidney Disease: Improving Global Outcomes* defines CKD as the presence of abnormalities of kidney structure or kidney function of any cause, that exist for at least 3 months[7,8]. It is associated with cardiovascular and all-cause mortality[9,10], and it may eventually progress to end-stage renal disease which requires renal replacement therapy (i.e. dialysis and kidney transplantation). CKD staging is based on risk classification of cardiovascular events and end-stage renal disease, and is currently determined by a combination of level of kidney function (assessed by estimated glomerular filtration rate, eGFR) and kidney damage (assessed by albuminuria) (**Table 1**). It is estimated that CKD affects 11-13% of the global population[11]. The incidence of CKD is increasing. Extrapolating from current trends, it has been projected that 50% of the US population will eventually develop some stage of CKD during their lifetime[12]. As such, CKD poses a major burden on patients and global health resources.

Traditional cardiovascular risk factors such as older age, overweight, and smoking predispose to CKD[13,14] but only explain a relatively small percentage of CKD cases. The most important risk factors for CKD are diabetes and hypertension, which together explain 50-70% of cases. However, it has been observed that CKD can

also occur in the absence of diabetes and hypertension[15]. Thus, a large proportion of CKD cases remains unexplained, warranting the identification of additional, non-traditional risk factors.

Table 1. Prognosis of CKD by categories of GFR and albuminuria. CKD is defined as abnormalities of kidney structure or function present >3 months, decreased GFR <60mL/min/1.73m² (G3) and/or at least moderately increased albuminuria (A2). Darker coloring indicates higher risk of cardiovascular events and end-stage renal disease. Adapted from KDIGO 2012.

					Persistent Albuminuria Categories		
					Description and range		
					A1	A2	A3
					Normal to mildy increased	Moderately increased	Severely increased
					<30 mg/g	30-300 mg/g	>300 mg/g
					≤3 mg/mmol	3-30 mg/mmol	>30 mg/mmol
GFR Categories (mL/min/1.73m²)	Description and Range	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			

Socioeconomic disparities in risk of CKD

Socioeconomic status, also referred to as socioeconomic position or social class, represents one's access to social and economic assets and resources[16].

CKD is unequally distributed across socioeconomic groups. Higher prevalence and incidence rates of CKD and end-stage renal disease have consistently been observed among those with low socioeconomic status, and socioeconomic gradients have been observed for the CKD markers eGFR and albuminuria[4,6]. It is not fully understood what drives the association between socioeconomic status and CKD, and little has therefore been achieved in reducing socioeconomic disparities in CKD. The limited understanding of the association may in part be due to differences within and between populations, reflected by the substantial between-study heterogeneity that has been observed in meta-analysis of the association. This may be explained by differences in CKD prevalence, ethnic composition, health behavior, prevalence of risk factors, and healthcare systems[17] between populations. Therefore, country and/or population-specific estimates of the relation should be made.

Some of the observed heterogeneity may also be explained by the socioeconomic indicator that is used. In health research, socioeconomic status is commonly measured by education, income, occupational level, area/neighborhood deprivation or any combination of these[16,18-21]. The indicators are not interchangeable[18] and the choice of indicator may itself be a source of heterogeneity between studies. For example, some evidence exists that education, not income, is associated with CKD in the Netherlands, whereas in the United States, income is more strongly associated with CKD than education[22]. Educational level is sometimes the preferred indicator of socioeconomic status as it is easy to measure and yields a high response rate. It theoretically captures one's knowledge related assets and cognitive abilities. Formal education is usually completed in young adulthood and therefore reflects early life socioeconomic status[19,21]. One advantage of education as an indicator of socioeconomic status in CKD research is that, in contrast to income, it is not affected by reverse causality (i.e. disease causing low education) given that CKD usually presents at older age.

Low socioeconomic status is not likely to increase risk of CKD in a direct manner. Rather, it is proposed to affect CKD risk through a wide range of intermediate pathways, including social (neighborhood deprivation, health care affordability, health care access), psychological (e.g. depression, stress), behavioral (smoking, poor diet), and biological factors (inflammation, obesity, hypertension)[23-26]. However, these propositions are not supported by data as only one cross-sectional study formally examined the contribution of potential mediators to the socioeconomic status -CKD association in the US[27]. More study on the pathways underlying socioeconomic disparities in CKD is therefore needed. Understanding the mechanisms through which socioeconomically disadvantaged groups (e.g. those with a low educational level) show higher vulnerability to CKD may prove helpful in designing interventions to reduce socioeconomic disparities in CKD. Given the challenges of intervening on education itself, managing and/or modifying downstream effects of low education to prevent CKD in disadvantaged groups, may be a more promising approach.

Genetic underpinnings of CKD

There is strong evidence for a genetic component to CKD. It tends to aggregate in families[28-31]. Furthermore, heritability of kidney function markers, estimated from family and twin studies, range between 36 and 75%, i.e. 36-75% of variance in

kidney markers can be attributed to genetic factors[32,33], although there is paucity of data from community-based samples. With advances in high-throughput measurement platforms, it became feasible to scan the entire human genome for possible leads towards causal genes. Such scans, called genome-wide association studies (GWAS), have identified a number of common variants, or single nucleotide polymorphisms (SNPs) (See **Box 1**), associated with kidney-related traits such as glomerular filtration rate, kidney function decline, urinary albumin, serum creatinine, and serum urea, in populations of European and Asian ancestry[34-42]. GWAS thus far identified >50 SNPs associated with creatinine-estimated glomerular filtration rate (eGFRcrea) in populations of European ancestry[34-37,43,44]. The phenotypic variance explained by the combined SNPs is modest (~4%); much of the genetic factors therefore remain to be found. Through advances in methodology and ever-increasing sample sizes, as well as the analysis of alternative markers of kidney function such as serum urea, it can be expected that new variants will be discovered. These new variants will explain larger amounts of phenotypic variance in the population, which may eventually lead to improved risk stratification and a deeper understanding of the mechanisms underlying CKD.

Genetics applied to social and clinical epidemiology

Although individual effects of known genetic variants associated with kidney outcomes are small, it may be possible to use the information hidden within these

Box 1. Genome-wide association studies and single nucleotide polymorphisms

Traditional linkage studies were highly successful in identifying genetic mutations underlying Mendelian diseases and traits (i.e. those with a single underlying gene). However, linkage analysis has proven ineffective for complex, polygenic traits that do not follow Mendelian inheritance patterns, such as height and blood pressure, and diseases such as diabetes. The development of high-throughput microarrays enabled researchers to scan the human genome for genetic markers associated with complex phenotypes. Such scans, known as genome-wide association studies (GWAS), typically involve the examination of millions of genetic markers called single nucleotide polymorphisms (SNPs). SNPs are variations in a single base pair, at a single location in the DNA sequence. SNPs located in the coding region of a gene may be synonymous (not affecting protein sequence) or non-synonymous (altering the amino acid sequence of protein). SNPs not in coding regions may tag causal genetic loci by association, or contribute to the disease or trait by affecting expression of genes.

variants to improve risk prediction of CKD in individuals as well as the population. For example, the effects of the 63 genetic variants associated with eGFRcrea may be aggregated into a genetic risk score, which holds promise as a reliable and accurate proxy for a genetic component to kidney function. For example, such a genetic risk score may be used to examine gene-environment interaction; recently it has been observed that higher socioeconomic status offsets genetic risk of obesity and diabetes[45,46], and it is possible that this also applies to genetic risk of CKD.

Furthermore, SNPs can be used as instrumental variables in a quasi-experimental design named Mendelian randomization[47,48]. This method exploits the random assortment and independent assignment of alleles to individuals. Analogous to a randomized clinical trial, individuals are randomly assigned to increased or decreased exposure to a risk factor based on their genotype. Due to the random assignment, confounding is minimized. Furthermore, given that the outcomes cannot influence one's genotype, reverse causation is unlikely. Therefore, under a number of assumptions, estimates of association derived from such Mendelian randomization analyses are considered causal estimates. This method is increasingly being applied to social and clinical epidemiology. For example, in recent Mendelian randomization studies, educational attainment has been implicated as a causal factor in smoking[49,50], obesity[51], and coronary heart disease[52]. These studies lend further support for a causal role of socioeconomic factors in disease risk. Given the large body of observational evidence on the socioeconomic status - CKD association, and that many of the underlying risk factors of coronary heart disease are similar to those of CKD, it is likely there is a causal role of socioeconomic factors in CKD risk as well.

Thesis outline

Aims

In this thesis, I aim to elucidate pathways leading to CKD in the general population. More specifically, in applying concepts from genetic epidemiology to social epidemiology, I hope to increase our understanding of socioeconomic disparities in CKD risk.

Research question 1

Is educational level associated with long-term risk of CKD in the general population? If so, what are mediators of this association? (**Chapter 2**)

Research question 2

Is low heart rate variability, an indicator of poor autonomic function, associated with increased risk of CKD in the general population? (**Chapter 3**)

Research question 3

CKD is observed to aggregate in families. What are the odds of developing CKD when a family member has CKD? What is the contribution of genetic factors to the CKD defining traits, eGFR and albuminuria, in the general population? (**Chapter 4**)

Research question 4

GWAS identified 53 SNPs associated with eGFRcrea. Is a genetic risk score based on these SNPs an accurate genetic proxy of kidney function? If so, can such a genetic risk score be used for CKD risk prediction? (**Chapter 5**)

Research question 5

Serum urea is an alternative marker of kidney function. Which are the genes that influence serum urea? What function do these genes have? Can we, through these genes, gain insights into the physiology of serum urea and kidney function, and into the pathways leading to kidney disease? (**Chapter 6**)

Research question 6

Does lower education amplify the negative consequences of a higher genetic predisposition to CKD? (**Chapter 7**)

Research question 7

Can we obtain *causal* estimates of the inverse association between education and CKD using genetic proxies of educational attainment? (**Chapter 8**)

To address the research questions in this thesis, we leverage data from large samples of the general population. The two most important are the Prevention of REnal and Vascular ENd-stage Disease (PREVEND) Study and the Lifelines cohort study and Biobank. Furthermore, we apply summary data from large GWAS consortia such as the Chronic Kidney Disease Genetics (CKDGen) Consortium, and the Social Science and Genetics Association Consortium (SSGAC). Information on data sources and study design, by thesis chapter, is provided in **Table 2**. Details on these sources are described in the referred chapter.

Thesis structure

A general introduction of this thesis is provided in **Chapter 1**. Here, concepts, constructs, and hypotheses underlying this thesis are discussed, and an overview of the available literature is provided. In **Chapter 2**, socioeconomic disparities, assessed by educational level, in long-term risk of CKD are examined. Furthermore, I explore potential underlying mechanisms of this association. In **Chapter 3**, I investigate the association of heart rate variability (HRV), a marker of poor autonomic function, with CKD. In **Chapter 4**, I construct a genetic risk score comprised of genetic variants associated with creatinine-estimated glomerular filtration rate. I then examine its cross-sectional and longitudinal associations with a number of complementary kidney outcomes to ascertain whether it is an accurate and clinically applicable representation of the genetics underlying kidney function. In **Chapter 5**, I describe a meta-analysis of GWAS to identify genetic variants associated with urea, an alternative marker of kidney function, in populations of European ancestry. In follow-up analyses, we attempt to characterize these variants and their relevance to urea physiology and kidney function and disease. In **Chapter 6**, I examine the familial aggregation of CKD, and estimate the relative contribution of genetic factors in CKD related traits. In **Chapter 7**, I address the question whether high socioeconomic status offsets genetic predisposition to reduced kidney function by examining the statistical interactions between education and a genetic risk score. In **Chapter 8**, I perform a Mendelian Randomization study to obtain causal estimates of the relation between education and kidney outcomes. Finally, in **Chapter 9**, I discuss the most important findings and their implications for clinical practice, research practice, and public health.

Table 2. Overview of thesis chapters: data sources, design, number of participants, determinants, and outcomes

Chapter		Data source(s)	Design	N	Determinants	Main outcome	Secondary outcome(s)
1	**General introduction**	-	-	-	-	-	-
2	**Educational level and risk of chronic kidney disease: longitudinal data from the PREVEND Study**	PREVEND	Cohort study	6,078	Educational level	CKD	eGFR
3	**Heart rate variability and its association with chronic kidney disease: results from the PREVEND Study**	PREVEND	Cohort study	4,605	Heart rate variability	CKD	eGFR
4	**Familial aggregation of chronic kidney disease and heritability of related biomarkers in the general population: The Lifelines Cohort Study**	Lifelines	Family study	155,936	-	CKD	eGFR, albuminuria, serum urea, uric acid, serum electrolytes
5	**Evaluation of a genetic risk score based on creatinine estimated glomerular filtration rate and its association with kidney outcomes**	PREVEND	Cohort study	3,649	Genetic risk score based on 53 eGFRcrea SNPs	CKD	eGFR albuminuria
6	**Genome-wide scan of serum urea in European populations identifies two novel loci**	Lifelines PREVEND NESDA EGCUT I+II	Two-stage genome-wide association study	20,391	Hypothesis-free: >2.5 x10^6 SNPs	serum urea	eGFR
7	**Interaction of educational level and genetic factors in kidney function**	PREVEND	Cohort study	3,597	Educational level, Genetic risk score based on 63 eGFRcrea SNPs	eGFR	-
8	**Causal effects of educational attainment on kidney outcomes: a Mendelian Randomization study**	SSGAC CKDGen Lifelines	Two-sample Mendelian Randomization study	>10^6	1271 SNPs for years of schooling	eGFR albuminuria	-
9	**General discussion**	-	-	-	-	-	-

REFERENCES

1. Berkman LF, Kawachi I, Glymour MM. *Social epidemiology.* Oxford University Press; 2014.
2. Mackenbach JP. *Health inequalities: Europe in profile.* Produced by COI for the Department of Health; 2006.
3. Mackenbach JP, Stirbu I, Roskam AR, et al. Socioeconomic inequalities in health in 22 european countries. *N Engl J Med*. 2008;358(23):2468-2481.
4. Vart P, Gansevoort RT, Joosten MM, Bultmann U, Reijneveld SA. Socioeconomic disparities in chronic kidney disease: A systematic review and meta-analysis. *Am J Prev Med*. 2015;48(5):580-592.
5. Vart P, van Zon SKR, Gansevoort RT, Bultmann U, Reijneveld SA. SES, chronic kidney disease, and race in the U.S.: A systematic review and meta-analysis. *Am J Prev Med*. 2017;53(5):730-739.
6. Zeng X, Liu J, Tao S, Hong HG, Li Y, Fu P. Associations between socioeconomic status and chronic kidney disease: A meta-analysis. *J Epidemiol Community Health*. 2018;72(4):270-279.
7. Levey AS, de Jong PE, Coresh J, et al. The definition, classification, and prognosis of chronic kidney disease: A KDIGO controversies conference report. *Kidney Int*. 2011;80(1):17-28.
8. Stevens PE, Levin A, Kidney Disease: Improving Global Outcomes Chronic Kidney Disease Guideline Development Work Group Members. 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):825-830.
9. Go AS, Chertow GM, Fan D, McCulloch CE, Hsu CY. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *N Engl J Med*. 2004;351(13):1296-1305.
10. Gansevoort RT, Correa-Rotter R, Hemmelgarn BR, et al. Chronic kidney disease and cardiovascular risk: Epidemiology, mechanisms, and prevention. *The Lancet*. 2013;382(9889):339-352.
11. Hill NR, Fatoba ST, Oke JL, et al. Global prevalence of chronic kidney disease–a systematic review and meta-analysis. *PloS one*. 2016;11(7):e0158765.
12. Grams ME, Chow EKH, Segev DL, Coresh J. Lifetime incidence of CKD stages 3-5 in the United States. *Am J Kidney Dis*. 2013;62(2):245-252.
13. Fox CS, Larson MG, Leip EP, Culleton B, Wilson PW, Levy D. Predictors of new-onset kidney disease in a community-based population. *JAMA*. 2004;291(7):844-850.
14. Muntner P, He J, Astor BC, Folsom AR, Coresh J. Traditional and nontraditional risk factors predict coronary heart disease in chronic kidney disease: Results from the Atherosclerosis Risk In Communities study. *J Am Soc Nephrol*. 2005;16(2):529-538.
15. 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):180-188.
16. Krieger N, Williams DR, Moss NE. Measuring social class in US public health research: Concepts, methodologies, and guidelines. *Annu Rev Public Health*. 1997;18(1):341-378.
17. Bruck K, Stel VS, Gambaro G, et al. CKD prevalence varies across the European general population. *J Am Soc Nephrol*. 2016;27(7):2135-2147.
18. Braveman PA, Cubbin C, Egerter S, et al. Socioeconomic status in health research: One size does not fit all. *JAMA*. 2005;294(22):2879-2888.
19. Galobardes B, Shaw M, Lawlor DA, Lynch JW, Davey Smith G. Indicators of socioeconomic position (part 1). *J Epidemiol Community Health*. 2006;60(1):7-12.
20. Galobardes B, Shaw M, Lawlor DA, Lynch JW, Davey Smith G. Indicators of socioeconomic position (part 2). *J Epidemiol Community Health*. 2006;60(2):95-101.
21. Galobardes B, Lynch J, Smith GD. Measuring socioeconomic position in health research. *Br Med Bull*. 2007;81(1):21.

22. Vart P, Gansevoort RT, Coresh J, Reijneveld SA, Bultmann U. Socioeconomic measures and CKD in the United States and the Netherlands. *Clin J Am Soc Nephrol*. 2013;8(10):1685-1693.

23. Adler NE, Newman K. Socioeconomic disparities in health: Pathways and policies. *Health Aff*. 2002;21(2):60-76.

24. Chen E, Miller GE. Socioeconomic status and health: Mediating and moderating factors. *Annu Rev Clin Psychol*. 2013;9:723-749.

25. Petrovic D, de Mestral C, Bochud M, et al. The contribution of health behaviors to socioeconomic inequalities in health: A systematic review. *Prev Med*. 2018;113:15-31.

26. Patzer RE, McClellan WM. Influence of race, ethnicity and socioeconomic status on kidney disease. *Nat Rev Nephrol*. 2012;8(9):533.

27. Vart P, Gansevoort RT, Crews DC, Reijneveld SA, Bultmann U. Mediators of the association between low socioeconomic status and chronic kidney disease in the United States. *Am J Epidemiol*. 2015;181(6):385-396.

28. Lei HH, Perneger TV, Klag MJ, Whelton PK, Coresh J. Familial aggregation of renal disease in a population-based case-control study. *J Am Soc Nephrol*. 1998;9(7):1270-1276.

29. Queiroz Madeira EP, da Rosa Santos O, Ferreira Santos SF, Alonso da Silva L, MacIntyre Innocenzi A, Santoro-Lopes G. Familial aggregation of end-stage kidney disease in Brazil. *Nephron*. 2002;91(4):666-670.

30. Satko SG, Sedor JR, Iyengar SK, Freedman BI. Familial clustering of chronic kidney disease. *Semin Dial*. 2007;20(3):229-236.

31. Skrunes R, Svarstad E, Reisaeter AV, Vikse BE. Familial clustering of ESRD in the Norwegian population. *Clin J Am Soc Nephrol*. 2014;9(10):1692-1700.

32. Satko SG, Freedman BI. The familial clustering of renal disease and related phenotypes. *Med Clin North Am*. 2005;89(3):447-456.

33. O'Seaghdha CM, Fox CS. Genome-wide association studies of chronic kidney disease: What have we learned? *Nat Rev Nephrol*. 2011;8(2):89-99.

34. Kottgen A, Glazer NL, Dehghan A, et al. Multiple loci associated with indices of renal function and chronic kidney disease. *Nat Genet*. 2009;41(6):712-717.

35. Kottgen A, Pattaro C, Boger CA, et al. New loci associated with kidney function and chronic kidney disease. *Nat Genet*. 2010;42(5):376-384.

36. Pattaro C, Teumer A, Gorski M, et al. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat Commun*. 2016;7.

37. Gorski M, van der Most PJ, Teumer A, et al. 1000 genomes-based meta-analysis identifies 10 novel loci for kidney function. *Scientific Reports*. 2017;7:45040.

38. Boger CA, Chen MH, Tin A, et al. CUBN is a gene locus for albuminuria. *J Am Soc Nephrol*. 2011;22(3):555-570.

39. Teumer A, Tin A, Sorice R, et al. Genome-wide association studies identify genetic loci associated with albuminuria in diabetes. *Diabetes*. 2016;65(3):803-817.

40. Lee J, Lee Y, Park B, Won S, Han JS, Heo NJ. Genome-wide association analysis identifies multiple loci associated with kidney disease-related traits in korean populations. *PLOS ONE*. 2018;13(3):e0194044.

41. Teumer A, Li Y, Ghasemi S, et al. Genome-wide association meta-analyses and fine-mapping elucidate pathways influencing albuminuria. *Nat Commun*. 2019;10(1):1-19.

42. Okada Y, Sim X, Go MJ, et al. Meta-analysis identifies multiple loci associated with kidney function-related traits in east asian populations. *Nat Genet*. 2012;44(8):904-909.

43. Chambers JC, Zhang W, Lord GM, et al. Genetic loci influencing kidney function and chronic kidney disease. *Nat Genet*. 2010;42(5):373-375.

44. Wuttke M, Li Y, Li M, et al. A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat Genet*. 2019;51(6):957-972.

45. Liu SY, Walter S, Marden J, et al. Genetic vulnerability to diabetes and obesity: Does education offset the risk? *Soc Sci Med*. 2015;127:150-158.

46. van Zon SKR, Reijneveld SA, van der Most PJ, Swertz MA, Bultmann U, Snieder H. The interaction of genetic predisposition and socioeconomic position with type 2 diabetes mellitus: Cross-sectional and longitudinal analyses from the lifelines cohort and biobank study. *Psychosom Med*. 2018;80(3):252-262.

47. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stat Med*. 2008;27(8):1133-1163.

48. Sekula P, Del Greco MF, Pattaro C, Kottgen A. Mendelian randomization as an approach to assess causality using observational data. *J Am Soc Nephrol*. 2016;27(11):3253-3265.

49. Gage SH, Bowden J, Davey Smith G, Munafo MR. Investigating causality in associations between education and smoking: A two-sample Mendelian randomization study. *Int J Epidemiol*. 2018;47(4):1131-1140.

50. Sanderson E, Davey Smith G, Bowden J, MunafÃ² MR. Mendelian randomisation analysis of the effect of educational attainment and cognitive ability on smoking behaviour. *Nat Commun*. 2019;10(1):2949.

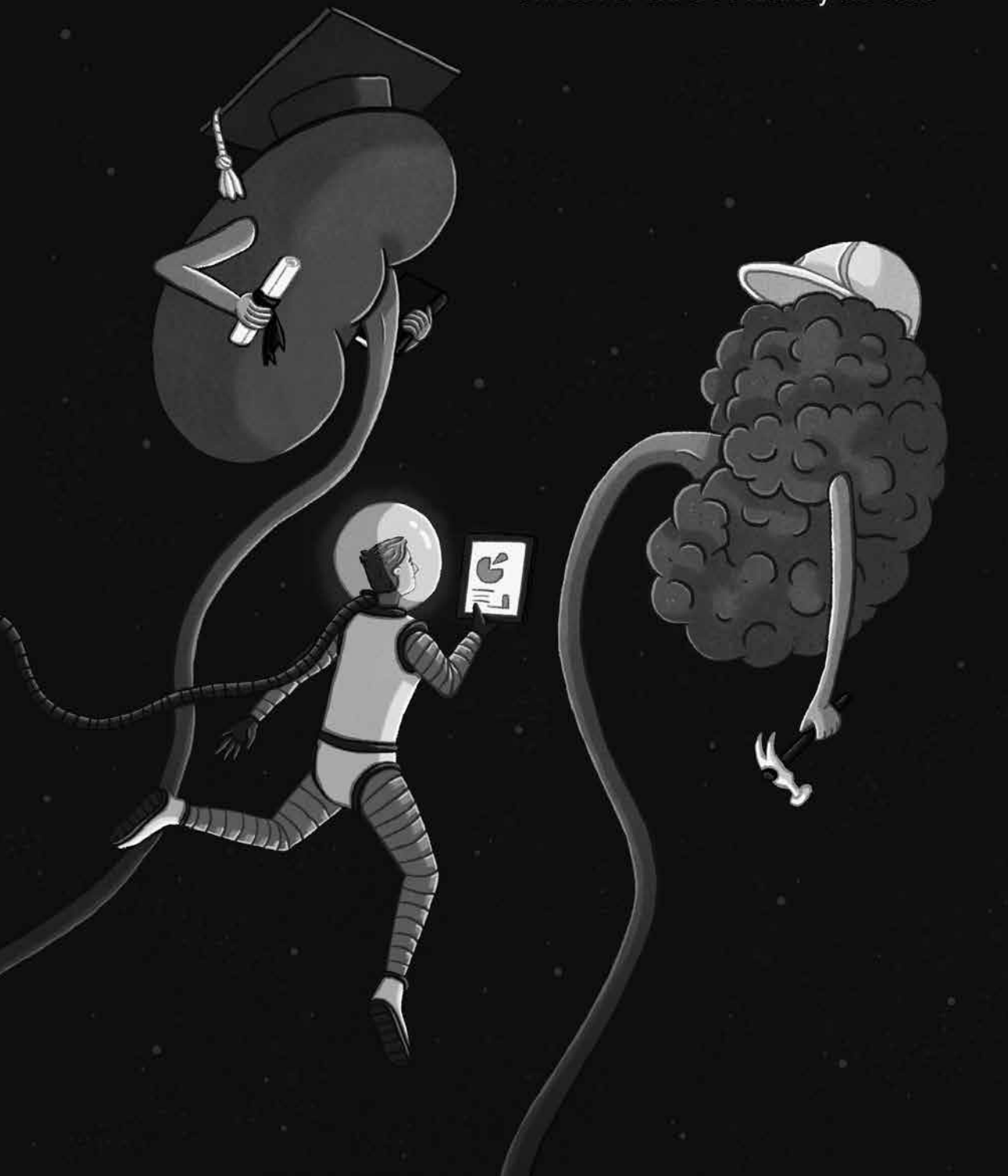
51. Böckerman P, Viinikainen J, Pulkki-Råback L, et al. Does higher education protect against obesity? Evidence using Mendelian randomization. *Prev Med*. 2017;101:195-198.

52. Tillmann T, Vaucher J, Okbay A, et al. Education and coronary heart disease: Mendelian randomisation study. *BMJ*. 2017;358:j3542.

1

PART I

Evaluating the effect of socioeconomic status and autonomic dysfunction on risk of chronic kidney disease



Educational level and risk of chronic kidney disease: Longitudinal data from the PREVEND study

Chris HL Thio, Priya Vart, Lyanne M Kieneker, Harold Snieder, Ron T Gansevoort, Ute Bültmann

Nephrol Dial Transplant 2018 (in press)

CHAPTER 2

ABSTRACT

Introduction. The longitudinal association between low education and chronic kidney disease (CKD) and its underlying mechanisms are poorly characterized. We therefore examined the association of low education with incident CKD and change in kidney function, and explored potential mediators of this association.

Methods. We analyzed data on 6078 participants from the community-based PREVEND Study. Educational level was categorized into low, medium, and high (<secondary, secondary/equivalent, >secondary schooling). Kidney function was assessed by estimating glomerular filtration rate (eGFR) by serum creatinine and cystatin C at five examinations during ~11 years of follow-up. Incident CKD was defined as new-onset eGFR<60mL/min/1.73m^2 and/or urinary albumin≥30mg/24h in those free of CKD at baseline. We estimated main effects with Cox regression and linear mixed models. In exploratory causal mediation analyses, we examined mediation by several potential risk factors.

Results. Incident CKD was observed in 861 (17%) participants. Lower education was associated with higher rates of incident CKD (low vs high education; HR[95%CI]=1.25 [1.05 to 1.48], p_{trend}=0.009) and accelerated eGFR decline (B[95%CI]=-0.15 [-0.21 to -0.09] mL/min/1.73m^2 per year, p_{trend}<0.001). The association between education and incident CKD was mediated by smoking, potassium excretion, BMI, WHR, and hypertension. Analysis on annual eGFR change in addition suggested mediation by magnesium excretion, protein intake, and diabetes.

Conclusions. In the general population, we observed an inverse association of educational level with CKD. Diabetes, and the modifiable risk factors smoking, poor diet, BMI, WHR, and hypertension are suggested to underlie this association. These findings provide support for targeted preventive policies to reduce socioeconomic disparities in kidney disease.

Keywords: chronic kidney disease, educational level, socioeconomic status, health disparities

ABBREVIATIONS

BMI = body-mass index
CKD = chronic kidney disease
eGFR = estimated glomerular filtration rate
PREVEND = Study Prevention of renal and vascular end-stage disease study
SES = socioeconomic status
UAE = urinary albumin excretion
WHR = waist-to-hip ratio

Supplementary material available at www.doi.org/10.1093/ndt/gfy361

BACKGROUND

Chronic kidney disease (CKD) is a heterogeneous group of disorders characterized by sustained diminished kidney function and/or kidney damage. CKD affects ~10-15% of the global population, and its incidence is increasing[1-4]. CKD can progress to end-stage renal disease (ESRD), and is associated with an increased incidence of cardiovascular disease and all-cause mortality[5,6]. As such, CKD poses a major burden on patients and global health resources.

CKD is unequally distributed across socioeconomic groups: higher prevalence and incidence rates of CKD and ESRD have consistently been observed among those with lower socioeconomic status (SES). Socioeconomic gradients have also been observed for eGFR and urinary albumin. However, large heterogeneity exists between studies of the SES-CKD association[7,8]. One possible explanation for this heterogeneity is that factors underlying the SES-CKD association vary between populations due to differences in e.g. ethnicity, lifestyle, prevalence of comorbid conditions, or healthcare[9,10]. Currently, the available literature is limited: 1) most observations were made in US-based cross-sectional data[7,8] and 2) European studies established cross-sectional associations of SES measures with CKD[11-13]; however no European study explicitly examined the association of SES with CKD, or mediators of this association, in a longitudinal setting. Hence, it is uncertain to what extent SES conveys risk of CKD in the European general population, and which factors underlie this association. Characterization of underlying mechanisms may help identify targets for disease prevention and management, thus help alleviate the burden of CKD and its consequences among disadvantaged populations.
Our aim was therefore to examine the strength of the association of SES with the longitudinal outcomes, CKD incidence and annual change in eGFR, in a sample of the Dutch general population. Furthermore, we explored health-related behaviors and comorbid conditions that potentially mediate this association.

MATERIALS AND METHODS

Study design and population

We used data from the Prevention of REnal and Vascular ENdstage Disease (PREVEND) cohort study. PREVEND was initiated to investigate the natural course of increased urinary albumin levels and its association to renal and vascular outcomes. Details of this study have been described elsewhere[14]. Briefly, 8592 individuals, sampled from the general population of Groningen,

the Netherlands, underwent extensive examination between 1997-1998. Four follow-up examinations were completed in 2003, 2006, 2008, and 2012. All subjects gave written informed consent. PREVEND was approved by the medical ethics committee of the University Medical Center Groningen and conducted in accordance with the Helsinki Declaration guidelines. In the present study, we excluded participants with incomplete data on educational level, kidney outcomes, or important covariates.

Measures

We defined CKD according to Kidney Disease: Improving Global Outcomes guidelines (eGFR<60mL/min/1.73m^2 or UAE≥30mg/24h)[15]. Incident cases were those participants free of CKD at baseline who developed CKD during follow-up. We calculated eGFR from serum creatinine and serum cystatin C, using the corresponding CKD-EPI equation[16].

Collection procedures of blood and two consecutive 24h-urine specimens at each examination has been described previously[17]. Measurement of serum creatinine was performed by an enzymatic method on a Roche Modular analyzer using reagents and calibrators from Roche (Roche Diagnostics, Mannheim, Germany), with intra- and interassay coefficients of variation of 0.9% and 2.9%, respectively. Serum cystatin C concentration was measured by a Gentian cystatin C Immunoassay (Gentian AS Moss, Norway) on a Modular analyzer (Roche Diagnostics). Cystatin C was calibrated directly using the standard supplied by the manufacturer (traceable to the International Federation of Clinical Chemistry Working Group for Standardization of Serum Cystatin C)[18]. Intra- and interassay coefficients of variation were <4.1% and <3.3%, respectively. Urinary albumin concentration (UAC) was measured by nephelometry with a lower threshold of detection of 2.3mg/L, and intra- and interassay coefficient of variation of 2.2% and 2.6%, respectively (Dade Behring Diagnostic, Marburg, Germany). UAC was multiplied by urine volume to obtain a value of UAE in mg/24h. The two 24h-urinary albumin values of each subject per examination were averaged.

SES was measured by educational level, categorized into low (no, primary, basic vocational, and secondary education), medium (senior secondary vocational and general senior secondary education), and high (higher professional and higher academic education) according to the International Standard Classification of

Education[19]. Furthermore, we examined associations of income as alternative measure of SES. For this, we categorized income into low, medium, and high according to tertiles of the ratio between reported income and the 1998 poverty line (1658 guilders per month).

Age, sex, and baseline eGFR were included as potential confounders. Included as potential mediators were: current smoking (self-reported yes/no), alcohol consumption (labelled as none, occasional [<10g/wk), light (10-69.9g/wk), moderate (70-210g/wk), heavier (>210g/wk)], 24h urinary excretions of sodium (Na^+), potassium (K^+), and magnesium (Mg^{2+}) (as surrogates for dietary intake of sodium, potassium, and magnesium), 24h protein intake (estimated from 24h urea excretion by the Maroni formula[20,21]), body-mass index (BMI, weight/height^2), waist-to-hip ratio (WHR, waist/hip circumference), diabetes (fasting glucose>7.0mmol/L, non-fasting glucose>11.0mmol/L, anti-diabetic treatment, or self-reported), hypertension (systolic blood pressure>140mmHg, diastolic blood pressure>90mmHg, blood pressure lowering treatment, or self-reported), hypercholesterolemia (total cholesterol≥6.21mmol/L, lipid lowering treatment, or self-reported). Covariates were collected at baseline by questionnaires, anthropometry, urine collections, or pharmacy records. Urinary concentrations of Na^+, K^+, and Mg^{2+} were determined as previously described[17,22].

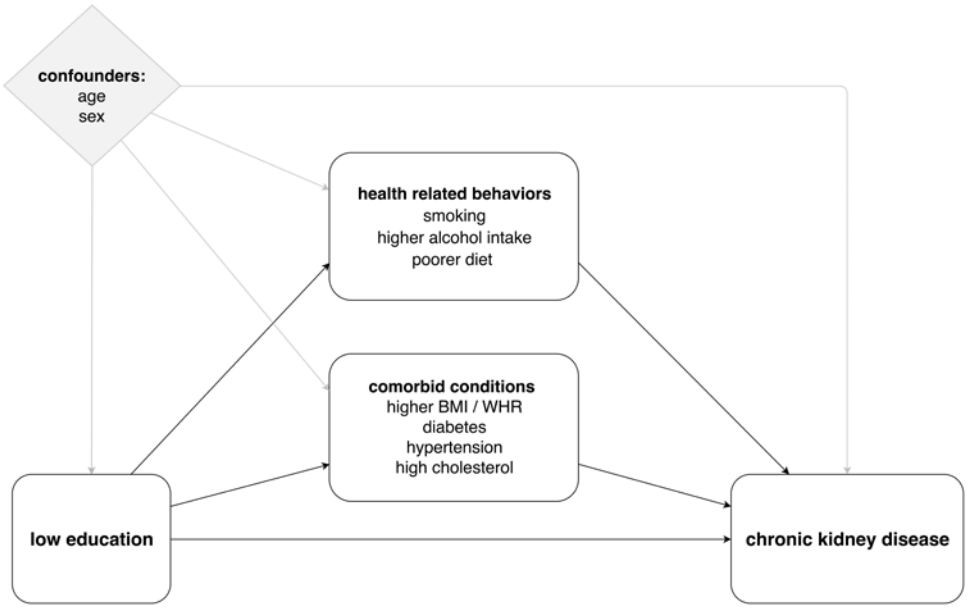
Statistical analyses

Statistical analyses were performed using R v3.4.1[23] and SPSS v23 software (IBM corp, Amonk, NY, USA) during years 2017 and 2018. Two-sided significance level was set at =0.05 unless otherwise stated. Baseline characteristics were examined for the total population and compared across categories of education using one-way ANOVA, Jonckheere-Terpstra, or ^{2}-tests for linear trend. We used the *survival* R-package[24] for Cox proportional hazards modelling of time to CKD. Time of CKD was estimated using a midpoint imputation method. Crude effects were examined in an unadjusted model. Next, we adjusted for age, sex, and baseline eGFR. In a final model, we introduced potential mediators. We calculated p for linear trend by analyzing education as a continuous rather than an ordinal variable. Using the *lme4* R-package[25], we estimated eGFR change by modelling eGFR as a function of time in a random intercept, random coefficient linear mixed model. To examine the crude effect of SES on annual eGFR change, an interaction term between time and SES was introduced. Next, we adjusted for age and sex, as well as their

interaction with time. Finally, we introduced all potential mediators, and the interaction of each with time.

Next, we performed exploratory mediation analyses. **Figure 1** shows a graph of hypothesized pathways tested in the present study. Main effects of potential mediators on kidney outcomes were examined with Cox proportional hazards and linear mixed models adjusting for age, sex, and baseline eGFR. Next, we used the *mediation* R-package[26] to estimate mediation within the counterfactual framework described by Imai et al[27]. Here, we simplified our statistical models by using one contrast for education (low vs high education). Furthermore, we used individual eGFR slopes (extracted from a linear mixed-effects model) as outcome variable in mediation analysis of eGFR change. Finally, we used parametric survival models implemented in the *survival* R-package. Due to these alternative methods, effects may deviate slightly from those of our main effects analyses. Each potential mediator was analyzed separately, adjusting for age, sex, and baseline eGFR. Any significant SES x mediator interaction was controlled for in the mediation

Figure 1. Graph of tested pathways through which low education could potentially lead to chronic kidney disease. BMI, body-mass index; WHR, waist-to-hip ratio. Poorer diet: high in sodium, low in potassium, low in magnesium, high in protein. Black arrows indicate a posited causal pathway; grey arrows indicate potential confounding pathways



model. Non-parametric bootstrap CIs and p-values were estimated from 1000 simulations. One-sided hypotheses were tested to assess potential mediators (i.e. current smoking[28], higher alcohol consumption, higher Na^+ excretion[29], lower K^+ excretion[17], lower Mg^{2+} excretion[30], higher estimated protein intake[31], higher BMI[32], higher WHR[33], diabetes, hypertension, and hypercholesterolemia).

In secondary analyses, we examined associations of education with incident CKD_{eGFR} (eGFR<60 mL/min/1.73m²), incident CKD_{UAE} (UAE≥30mg/24h) and annual change in UAE (natural-log transformed to approximate normality, *ln*UAE).

RESULTS

Baseline characteristics by educational level for 6078 participants with complete baseline data are presented in **Table 1**. Traditional risk factors (i.e. diabetes, hypertension, high cholesterol, smoking, higher BMI) were more prevalent in participants with low education. At baseline, low education participants were more likely to have CKD, lower eGFR, and higher UAE compared to high education participants. A higher attrition rate was observed for participants with low education: follow-up time was shorter for these participants. Low education was univariably associated with lower dietary quality as indicated by higher Na^+ excretion, lower K^+ excretion, lower Mg^{2+} excretion, and higher protein intake. Low education participants reported less alcohol consumption.

After excluding N=883 participants with baseline CKD, N=5195 remained for time-to-CKD analysis. Among these, 861 (17%) experienced new-onset CKD, with a significant socioeconomic gradient (low; med; high education: 22%; 14%; 12%, ^{2}[df] =62.8[1], p_{trend}<0.001). In the crude model, we observed an inverse association of education with CKD, again with a significant gradient (low vs high education: HR [95%CI] =1.97 [1.67 to 2.32], p_{trend}<0.001; **Table 2**). After adjusting for age, sex, and baseline eGFR, the association was attenuated, but significance remained (low vs high education: HR [95%CI] =1.25 [1.05 to 1.48], p_{trend}=0.009). After introducing all potential mediators to the model, the education-CKD association was no longer significant, suggesting mediation within our hypothesized framework (**Figure 1**).

Average estimated annual eGFR change for the total N=6078 population was -0.93 (95%CI: -0.95 to -0.91) mL/min/1.73m² per year. Low education was associated with accelerated eGFR change, with a significant gradient (low vs high education: B

Table 1. Baseline characteristics by categories of educational level.

	Total	Educational level			p_{trend}
		Low (< secondary)	Medium (secondary or equivalent)	High (> secondary)	
N	6078	2637	1565	1876	
Males	3071 (51%)	1223 (46%)	855 (55%)	993 (53%)	**0.005**
Age, years	48 [39-59]	54 [45-63]	44 [36-54]	43 [37-51]	**<0.001**
BMI, kg/m²	26 (4.1)	27 (4.3)	26 (4.0)	25 (3.3)	**<0.001**
WHR	0.88 (0.09)	0.90 (0.09)	0.87 (0.09)	0.86 (0.09)	**<0.001**
Current smoking	1956 (32%)	951 (36%)	529 (34%)	476 (25%)	**<0.001**
Alcohol					
None	1438 (24%)	881 (33%)	346 (22%)	211 (11%)	
Occasional (<10 g/wk)	956 (16%)	436 (17%)	263 (17%)	257 (14%)	
Light (10-69.9 g/wk)	2120 (35%)	782 (30%)	573 (37%)	765 (41%)	**<0.001**
Moderate (70-210 g/wk)	1252 (21%)	404 (15%)	303 (19%)	545 (29%)	
Heavier (>210 g/wk)	312 (5%)	134 (5%)	80 (5%)	98 (5%)	
Na^+ excretion (mmol/24h)	143 (51)	143 (52)	145 (51)	140 (48)	**0.021**
K^+ excretion (mmol/24h)	72 (21)	69 (20)	73 (22)	76 (21)	**<0.001**
Mg^{2+} excretion (mmol/24h)	3.9 (1.5)	3.8 (1.5)	4.0 (1.6)	4.1 (1.5)	**<0.001**
Estimated protein intake (g/kg/24h)	1.16 (0.26)	1.18 (0.28)	1.15 (0.26)	1.16 (0.24)	**0.005**
Diabetes	202 (3%)	130 (5%)	43 (3%)	29 (2%)	**<0.001**
Hypertension	1912 (31%)	1112 (42%)	418 (27%)	382 (20%)	**<0.001**
High cholesterol	1879 (31%)	1062 (40%)	423 (27%)	4394 (21%)	**<0.001**
Creatinine (µmol/L)	72 (16)	72 (17)	72 (15)	73 (14)	**0.015**
Cystatin C (mg/L)	0.89 (0.17)	0.91 (0.19)	0.88 (0.16)	0.86 (0.14)	**<0.001**
eGFR, ml/min/1.73m²	95 (17)	91 (17)	98 (16)	99 (15)	**<0.001**
UAE, mg/24h	9.1 [6.3-16]	10 [6.3-20]	8.9 [6.2-15]	8.4 [6.2-13]	**<0.001**
CKD at baseline	883 (15%)	510 (19%)	199 (13%)	174 (9%)	**<0.001**
CKD_{eGFR} at baseline	167 (3%)	109 (4%)	37 (2%)	21 (1%)	**<0.001**
CKD_{UAE} at baseline	805 (13%)	457 (17%)	181 (12%)	167 (9%)	**<0.001**
Follow-up time, yrs	11.2 [8.5-12.1]	11.1 [7.0-11.8]	11.3 [9.3-12.2]	11.4 [10.6-12.4]	**<0.001**

Baseline characteristics by categories of educational level. Data is presented as mean (standard deviation), median (interquartile range), and number (%) where appropriate. P-values reflect significance of a linear trend across categories of educational level, using one-way ANOVA, χ^2, or Jonckheere-Terpstra tests where appropriate.

Abbreviations: BMI, body-mass index; WHR, waist-to-hip ratio; eGFR, estimated glomerular filtration rate; UAE, urinary albumin excretion; CKD, chronic kidney disease

[95%CI] =-0.15 [-0.21 to -0.09], p_{trend}<0.001, adjusted for age and sex; **Table 2**). Addition of potential mediators to the model attenuated the association, although significance remained (low vs high education: B [95%CI] =-0.11 [-0.16 to -0.04], p_{trend}<0.001).

All potential mediators were associated with either CKD or annual eGFR change (Supplementary Table S4). We tested interactions of education with each potential mediator separately (age, sex, and baseline eGFR adjusted); none were significant

Table 2. Association of education with incident CKD (Panel A) and annual change in eGFR (Panel B).

A) Incident CKD				
		Educational level		
	Low (<secondary)	**Medium (secondary/equivalent)**	**High (>secondary)**	
N=5195	**N=2127**	**N=1366**	**N=1702**	**p_{trend}**
Events N=861	460 (22%)	193 (14%)	208 (12%)	**<0.001**
	HR (95%CI)			
Model 1	1.97 (1.67 to 2.32)	1.17 (0.96 to 1.42)	(ref.)	**<0.001**
Model 2[a]	1.25 (1.05 to 1.48)	1.07 (0.88 to 1.30)	(ref.)	**0.009**
Model 3	1.02 (0.85 to 1.22)	0.97 (0.80 to 1.19)	(ref.)	0.789
B) Annual eGFR change				
		Educational level		
	Low (<secondary)	**Medium (secondary/equivalent)**	**High (>secondary)**	
N=6078	**N=2637**	**N=1565**	**N=1876**	**p_{trend}**
	B (95%CI)			
Model 1[b]	-0.30 (-0.36 to -0.24)	-0.10 (-0.17 to -0.04)	(ref.)	**<0.001**
Model 2[b]	-0.15 (-0.21 to -0.09)	-0.08 (-0.14 to -0.02)	(ref.)	**<0.001**
Model 3[b]	-0.11 (-0.16 to -0.04)	-0.06 (-0.12 to 0.00)	(ref.)	**<0.001**

Data are presented as hazard ratio (95%CI) or unstandardized regression coefficient (95%CI, in mL/min/1.73m^2 per year).
Abbreviations: CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate.
Model 1: Crude Educational level (high educational level is reference category)
Model 2: Model 1 + age, sex, [a](and in addition baseline eGFR), [b](and in addition their interaction with time)
Model 3: Model 2 + potential mediators (body-mass index, waist-to-hip ratio, smoking, alcohol use, Na^+ excretion, K^+ excretion, Mg^{2+} excretion, estimated protein intake, diabetes, hypertension, high cholesterol) [b](and in addition their interaction with time)

(p>0.05). The association of low education and CKD was mediated by higher likelihood of smoking (proportion mediated [95%CI] =0.14 [0.02 to 0.51), p=0.009; **Table 3**), lower 24h K^+-excretion (0.12 [0.02 to 0.45], p=0.008), higher BMI (0.29 [0.13 to 0.96], p=0.004), higher WHR (0.31 [0.14 to 1.16], p=0.002), and higher prevalence of hypertension (0.14 [0.05 to 0.47], p=0.006) in this subpopulation. We observed no significant mediation by 24h Na^+ excretion, Mg^{2+} excretion, protein intake, diabetes, or hypercholesterolemia.

There were no education x mediator interactions with eGFR change as outcome except for smoking (low education [vs high education] x smoking: B [95%CI] =-0.07 [-0.10 to -0.05], p=0.01). The association of low education and accelerated eGFR decline was mediated by lower 24h K^+ excretion (**Table 3**, proportion mediated [95%CI] =0.08 [0.03 to 0.16], one-sided p<0.001), higher BMI (0.22 [0.12 to 0.42], p<0.001), higher WHR (0.09 [0.01 to 0.18], p=0.008), and higher prevalence of

hypertension (0.13 [0.08 to 0.24], p<0.001). Additionally, lower Mg^{2+} excretion (0.03 [-0.003 to 0.07], p=0.030), higher protein intake (0.01 [-0.001 to 0.04], p=0.032) and higher prevalence of diabetes (0.04 [0.01 to 0.10], p=0.009) mediated the association of low education with accelerated eGFR decline. A protective effect of smoking on eGFR change was observed; higher prevalence of smoking in those with low education appeared to offset risk of accelerated eGFR decline (proportion mediated [95%CI] =-0.12 [-0.22 to -0.06], p=1.000). Higher alcohol consumption was not a mediating risk factor, rather, alcohol seemed protective of CKD and accelerated eGFR decline (Supplementary Table S4). Estimates of average causal mediation effects and direct effects are listed in Supplementary Tables S5-6.

No significant associations between education with CKD_{eGFR} or CKD_{UAE} were found, although directions of effect for these outcomes were consistent with our main analysis (Supplementary Table S1-S2). Average estimated increase in UAE for

Table 3. Mediators of the association between Educational level and kidney outcomes

	Incident CKD			**Annual change in eGFR**	
Mediators	**Proportion mediated** (95%CI)	**p**		**Proportion mediated** (95%CI)	**p**
Health-related behaviors					
Smoking	**0.14 (0.02 to 0.51)**	**0.009**		-0.12 (-0.23 to -0.05) [a]	1.000
Alcohol	0.24 (0.05 to 0.99)	0.989		0.26 (0.16 to 0.49)	1.000
24h Na^+ excretion	-0.01 (-0.09 to 0.09)	0.431		0.01 (-0.02 to 0.06)	0.216
24h K^+ excretion	**0.12 (0.02 to 0.45)**	**0.008**		**0.08 (0.03 to 0.16)**	**<0.001**
24h Mg^{2+} excretion	0.04 (-0.03 to 0.18)	0.098		**0.03 (-0.003 to 0.07)**	**0.030**
Estimated 24h protein intake	0.001 (-0.002 to 0.04)	0.447		**0.01 (-0.001 to 0.04)**	**0.032**
Comorbid conditions					
BMI	**0.29 (0.13 to 0.96)**	**0.004**		**0.22 (0.12 to 0.42)**	**<0.001**
WHR	**0.31 (0.14 to 1.16)**	**0.002**		**0.09 (0.01 to 0.18)**	**0.008**
Diabetes	0.08 (-0.005 to 0.06)	0.064		**0.04 (0.01 to 0.10)**	**0.009**
Hypertension	**0.14 (0.05 to 0.47)**	**0.006**		**0.13 (0.08 to 0.24)**	**<0.001**
Hypercholesterolemia	0.02 (-0.05 to 0.13)	0.223		-0.04 (-0.10 to 0.01)	0.962

Results from causal mediation analysis. N=6078. Effects are reported as proportion mediated of the association between education (low vs high) and kidney outcomes. Non-parametric bootstrap confidence intervals and one-sided p-values are estimated from 1000 simulations. Estimates are conditioned on age, sex, and baseline eGFR.

One-sided hypotheses were that low education leads to steeper eGFR decline through: current smoking, higher alcohol consumption, higher Na^+ excretion, lower K^+ excretion, lower Mg^{2+} excretion, higher protein intake, higher BMI, higher WHR, diabetes, hypertension, and hypercholesterolemia.

[a] In addition adjusted for the interaction term educational level x smoking.

the total population was 1.1% (95%CI: 0.9% to 1.3%) per year. Low education was associated with accelerated increase in UAE (low vs high education: 0.7% [0.2% to 0.11%] accelerated increase in UAE per year, p_{trend}=0.003), but no longer significantly after adjusting for age and sex (Supplementary Table S3). There were no significant associations of household income, as alternative measure of socioeconomic status, with kidney outcomes after confounder adjustment (data not shown).

DISCUSSION

In a middle-aged community-based cohort, we examined the associations of SES, as indicated by educational level, with the longitudinal kidney outcomes, incident CKD and eGFR decline. Low education was associated with higher incidence rates of CKD, independent of age, sex, and baseline eGFR, but not of potential mediators. Furthermore, low education was associated with accelerated eGFR decline, independent of age, sex, and potential mediators. Exploratory longitudinal mediation analysis suggested that the association between education and CKD can partly be explained by diabetes and the modifiable risk factors, BMI, WHR, smoking, potassium, and hypertension. No significant associations of household income with kidney outcomes were observed.

With this longitudinal study, we corroborate previous cross-sectional observations that in the Netherlands, education, not income, is associated with kidney outcomes[12]. Recent longitudinal data from the US-based Atherosclerosis Risk in Communities study show effects of education on CKD incidence and eGFR decline comparable to the present data[34]. However, in contrast to the Netherlands, income is associated to CKD in the US[12,34]. Possible explanations for this discrepancy are: 1) in the US, healthcare access is income-dependent[35], and 2) there is larger income inequality compared to the Netherlands[36].

Our results are generally consistent with a previous mediation analysis on the SES-CKD association. This study assessed SES by household income, and was performed in a cross-sectional sample of the general US population[37]. Similar to that study, we observed mediation by smoking, (abdominal) obesity, diabetes, and hypertension. However, we could not corroborate a mediation effect of hypercholesterolemia. Vart et al[37] used questionnaires on availability of fruits and vegetables at home to assess dietary quality but did not observe mediation. In contrast, we used urinary measures to objectively assess dietary intake of various nutrients. We found strong

mediation effects of lower potassium intake on both incident CKD and accelerated eGFR decline, as well as suggestions for effects of lower magnesium intake and higher protein intake.

Of all nutrients examined in the present study, lower potassium intake was the strongest mediator. A large body of epidemiological data shows that low SES is associated with poor diet, especially with a lower consumption of micronutrients such as potassium[38]. Low potassium intake was previously observed to associate with an increased risk of incident hypertension[39], but also with incident CKD independent of hypertension[17]. A proposed mechanism involves induction of tubulointerstitial injury by ammoniagenesis caused by potassium deficiency[40,41]. Furthermore, potassium itself might be renoprotective by upregulating renal kinins[42]. On the other hand, potassium intake might reflect dietary quality more generally. The main dietary sources of potassium are fruits/vegetables, legumes, whole grains, and dairy products[43]. These potassium-rich foods contain fibers, polyphenols, antioxidants, and vitamins, which have health benefits[44] that may be renoprotective.

Interestingly, no mediation through sodium intake was observed. High sodium intake reflects poor diet due to its high content in processed foods[45,46], and is associated with the major renal risk factor, hypertension[47]; we therefore expected sodium to mediate the relation between education and CKD. However, we did not observe a strong educational gradient in sodium at baseline (**Table 1**). Furthermore, sodium intake was not found to be associated with CKD in PREVEND[17], which likely explains the observed lack of mediation in the present study.

Three counterintuitive findings need to be addressed. Firstly, despite its association with an elevated risk of CKD (concordant with literature[28]), smoking was associated with decelerated eGFR decline. We therefore further examined the main effect of smoking on eGFR decline in fully adjusted models: compared to non-smokers, smokers had lower baseline eGFR, and despite decelerated decline, eGFR on average remained lower in these participants (data not shown). Therefore, this finding is likely the result of a floor effect. Secondly, alcohol consumption was inversely associated with risk of CKD and eGFR decline. Moreover, lower alcohol consumption among low education participants partly explained the elevated risk of CKD. This may be due to residual confounding, a sick quitter/sick non-starter effect[48], or a cohort-specific effect; for a detailed discussion we refer to a study

by Koning et al. that previously observed this association in PREVEND[49]. Thirdly, mediation effects of diabetes were significant on eGFR decline, but only borderline significant (p=0.064) on incident CKD. This is likely the result of reduced statistical power of a dichotomous outcome compared to a continuous outcome, and low prevalence of diabetes (3% at baseline) in the PREVEND sample.

The mechanisms underlying the education-CKD association are incompletely understood. We therefore tested several biologically plausible mediating pathways. However, some are overlapping (e.g. BMI, WHR), or on the same causal pathway (e.g. low potassium intake leading to CKD possibly through hypertension). Hence, we examined each mediator separately, correcting only for age, sex, and baseline eGFR to prevent overadjustment. Due to sparse adjustment and the observational nature of PREVEND, we cannot exclude residual confounding. However, results were broadly concordant with the literature, i.e. effects were generally in the hypothesized direction. Therefore, any confounding has likely only biased magnitude, not direction, of mediation effects. Future work may involve further characterization of the education-CKD association by estimating effects of multiple mediators relative to one another using multivariable techniques (e.g. structural equation modelling or the counterfactual approach described by Lange et al[50,51]).

To the best of our knowledge, the present study is the first in Europe examining the longitudinal association between education and CKD in the general population, and the first exploring its underlying mechanisms in a longitudinal setting. Strengths of this study are its considerable size (N=6078) and follow-up time (~11 years). GFR was estimated from serial measurements of serum creatinine and cystatin C, currently considered the best proxy of kidney function in population-based studies. Furthermore, data on urinary albumin was available for all included participants. Finally, dietary variables were objectively measured from 24h urinary collections. Several limitations should be addressed. Firstly, PREVEND consists of >95% whites; we therefore could not address the influence of ethnicity in the education-CKD association. Secondly, we observed a higher attrition rate of participants with low education, which may have resulted in a bias towards the null. Thirdly, we lacked baseline information on several potential mediators (e.g. physical activity/sedentary time, healthcare access, health literacy, psychological factors). Finally, only individual-level socioeconomic data were available; we therefore could not examine effects of area-level SES.
In an effort to characterize socioeconomic disparities in CKD, we explored a number

of plausible mediating pathways (i.e. health behaviors and clinical risk factors) that link education to CKD. Future research may focus on e.g. 1) confirming the pathways suggested in the present study; 2) exploring other potential mediating factors such as health care access, health literacy, and psychological factors; 3) establishing the interrelationship between these factors. Understanding how and why socioeconomically disadvantaged groups (e.g. those with a lower educational level) show higher vulnerability to CKD may prove helpful in designing interventions to reduce socioeconomic disparities in CKD. Given the challenges of intervening on education itself, managing and/or modifying downstream effects of low education may be a more promising approach.

To conclude: in the Dutch general population, low SES, as indicated by educational level, is associated with elevated risk of CKD. This association is suggested to be driven by higher rates of diabetes and the modifiable risk factors, (abdominal) obesity, smoking, low potassium intake, and hypertension, in those with lower education. The data presented are a first step towards potential targeted public health interventions to reduce socioeconomic health disparities.

ACKNOWLEDGEMENTS

The PREVEND Study in general was funded by the Dutch Kidney Foundation (grant E.033). The funding source had no role in study design; in collection, analysis, or interpretation of the data; in writing of the report; or in the decision to submit for publication. Results have been previously presented as an abstract at the ERA-EDTA conference in Copenhagen, Denmark, May 2018. CT and PV conceived and designed the study. RG contributed to data acquisition. All authors contributed to either analysis or interpretation of the data. CT and LK drafted the manuscript. PV, LK, HS, RG, UB revised the article. HS, RG, UB supervised the work. All authors approved of the final version of the manuscript.

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

1. Grams ME, Chow EKH, Segev DL, Coresh J. Lifetime incidence of CKD stages 3-5 in the united states. *Am J Kidney Dis*. 2013;62(2):245-252.

2. Zhang QL, Rothenbacher D. Prevalence of chronic kidney disease in population-based studies: Systematic review. *BMC Public Health*. 2008;8:117-2458-8-117.

3. Jha V, Garcia-Garcia G, Iseki K, et al. Chronic kidney disease: Global dimension and perspectives. *The Lancet*. 2013;382(9888):260-272.

4. El Nahas AM, Bello AK. Chronic kidney disease: The global challenge. *The Lancet*. 2005;365(9456):331-340.

5. Gansevoort RT, Correa-Rotter R, Hemmelgarn BR, et al. Chronic kidney disease and cardiovascular risk: Epidemiology, mechanisms, and prevention. *The Lancet*. 2013;382(9889):339-352.

6. Go AS, Chertow GM, Fan D, McCulloch CE, Hsu CY. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *N Engl J Med*. 2004;351(13):1296-1305.

7. Vart P, Gansevoort RT, Joosten MM, Bultmann U, Reijneveld SA. Socioeconomic disparities in chronic kidney disease: A systematic review and meta-analysis. *Am J Prev Med*. 2015;48(5):580-592.

8. Zeng X, Liu J, Tao S, Hong HG, Li Y, Fu P. Associations between socioeconomic status and chronic kidney disease: A meta-analysis. *J Epidemiol Community Health*. 2018;72(4):270-279.

9. Bruck K, Stel VS, Gambaro G, et al. CKD prevalence varies across the European general population. *J Am Soc Nephrol*. 2016;27(7):2135-2147.

10. Patzer RE, McClellan WM. Influence of race, ethnicity and socioeconomic status on kidney disease. *Nature Reviews Nephrology*. 2012;8(9):533.

11. Bello AK, Peters J, Rigby J, Rahman AA, El Nahas M. Socioeconomic status and chronic kidney disease at presentation to a renal service in the United Kingdom. *Clin J Am Soc Nephrol*. 2008;3(5):1316-1323.

12. Vart P, Gansevoort RT, Coresh J, Reijneveld SA, Bultmann U. Socioeconomic measures and CKD in the United States and the Netherlands. *Clin J Am Soc Nephrol*. 2013;8(10):1685-1693.

13. Chudek J, Wieczorowska-Tobis K, Zejda J, et al. The prevalence of chronic kidney disease and its relation to socioeconomic conditions in an elderly Polish population: Results from the national population-based study PolSenior. *Nephrol Dial Transplant*. 2013;29(5):1073-1082.

14. Pinto-Sietsma SJ, Janssen WM, Hillege HL, Navis G, De Zeeuw D, De Jong PE. Urinary albumin excretion is associated with renal functional abnormalities in a nondiabetic population. *J Am Soc Nephrol*. 2000;11(10):1882-1888.

15. Levey AS, de Jong PE, Coresh J, et al. The definition, classification, and prognosis of chronic kidney disease: A KDIGO controversies conference report. *Kidney Int*. 2011;80(1):17-28.

16. Inker LA, Schmid CH, Tighiouart H, et al. Estimating glomerular filtration rate from serum creatinine and cystatin C. *N Engl J Med*. 2012;367(1):20-29.

17. Kieneker LM, Bakker SJ, de Boer RA, Navis GJ, Gansevoort RT, Joosten MM. Low potassium excretion but not high sodium excretion is associated with increased risk of developing chronic kidney disease. *Kidney Int*. 2016;90(4):888-896.

18. Grubb A, Blirup-Jensen S, Lindstrom V, et al. First certified reference material for cystatin C in human serum ERM-DA471/IFCC. *Clin Chem Lab Med*. 2010;48(11):1619-1621.

19. UNESCO Institute for Statistics. International standard classification of education. . 2011. Maroni BJ, Steinman TI, Mitch WE. A method for estimating nitrogen intake of patients with chronic renal failure. *Kidney Int*. 1985;27(1):58-65.

20. Halbesma N, Bakker SJ, Jansen DF, et al. High protein intake associates with cardiovascular events but not with loss of renal function. *J Am Soc Nephrol*. 2009;20(8):1797-1804.

21. Joosten MM, Gansevoort RT, Mukamal KJ, et al. Urinary magnesium excretion and risk of hypertension: The prevention of renal and vascular end-stage disease study. *Hypertension*. 2013;61(6):1161-1167.

22. R Core Team. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. www.R-project.org/. 2014.

23. Therneau T. A package for survival analysis in S. R package version 2.38. *Retrieved from www.CRAN.R-project.org/package= survival*. 2015.

24. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*. 2014.

25. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. Mediation: R package for causal mediation analysis. 2014.

26. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. *Psychol Methods*. 2010;15(4):309.

27. Xia J, Wang L, Ma Z, et al. Cigarette smoking and chronic kidney disease in the general population: A systematic review and meta-analysis of prospective cohort studies. *Nephrology Dialysis Transplantation*. 2017;32(3):475-487.

28. Smyth A, O'donnell MJ, Yusuf S, et al. Sodium intake and renal outcomes: A systematic review. *Am J Hypertension*. 2014;27(10):1277-1284.

29. Rebholz CM, Tin A, Liu Y, et al. Dietary magnesium and kidney function decline: The healthy aging in neighborhoods of diversity across the life span study. *Am J Nephrol*. 2016;44(5):381-387.

30. Malhotra R, Lipworth L, Cavanaugh KL, et al. Protein intake and long-term change in glomerular filtration rate in the jackson heart study. *J Renal Nutrition*. 2018.

31. Herrington WG, Smith M, Bankhead C, et al. Body-mass index and risk of advanced chronic kidney disease: Prospective analyses from a primary care cohort of 1.4 million adults in England. *PloS one*. 2017;12(3):e0173515.

32. Elsayed EF, Sarnak MJ, Tighiouart H, et al. Waist-to-hip ratio, body mass index, and subsequent kidney disease and death. *Am J Kidney Dis*. 2008;52(1):29-38.

33. Vart P, Grams ME, Ballew SH, Woodward M, Coresh J, Matsushita K. Socioeconomic status and risk of kidney dysfunction: The Atherosclerosis Risk in Communities study. *Nephrol Dial Transplant*. 2018.

34. Agrawal V, Jaar BG, Frisby XY, et al. Access to health care among adults evaluated for CKD: Findings from the kidney early evaluation program (KEEP). *Am J Kidney Dis*. 2012;59(3):S5-S15.

35. Zheng H. Do people die from income inequality of a decade ago? *Soc Sci Med*. 2012;75(1):36-45.

36. Vart P, Gansevoort RT, Crews DC, Reijneveld SA, Bultmann U. Mediators of the association between low socioeconomic status and chronic kidney disease in the United States. *Am J Epidemiol*. 2015;181(6):385-396.

37. Darmon N, Drewnowski A. Does social class predict diet quality?. *Am J Clin Nutr*. 2008;87(5):1107-1117.

38. Kieneker LM, Gansevoort RT, Mukamal KJ, et al. Urinary potassium excretion and risk of developing hypertension: The prevention of renal and vascular end-stage disease study. *Hypertension*. 2014;64(4):769-776.

39. Nath KA, Hostetter MK, Hostetter TH. Increased ammoniagenesis as a determinant of progressive renal injury. *Am J Kidney Dis*. 1991;17(6):654-657.

40. Tolins JP, Hostetter MK, Hostetter TH. Hypokalemic nephropathy in the rat. role of ammonia in chronic tubular injury. *J Clin Invest*. 1987;79(5):1447-1458.

41. Ardiles L, Cardenas A, Burgos ME, et al. Antihypertensive and renoprotective effect of the kinin pathway activated by potassium in a model of salt sensitivity following overload proteinuria. *Am J Physiol-Renal Physiol*. 2013;304(12):F1399-F1410.

42. Hendriksen M, van Rossum C. Kalium inname: Risico van hyperkaliëmie?: Overzicht van beschikbare gegevens in Nederland. 2015.

43. Slavin JL, Lloyd B. Health benefits of fruits and vegetables. *Adv Nutr*. 2012;3(4):506-516.

44. Anderson CA, Appel LJ, Okuda N, et al. Dietary sources of sodium in China, Japan, the United Kingdom, and the United States, women and men aged 40 to 59 years: The INTERMAP study. *J Am Diet Assoc*. 2010;110(5):736-745.

45. Van Rossum C, Fransen H, Verkaik-Kloosterman J, Buurma-Rethans E, Ocké M. Dutch national food consumption survey 2007-2010: Diet of children and adults aged 7 to 69 years. 2011.

46. Aburto NJ, Ziolkovska A, Hooper L, Elliott P, Cappuccio FP, Meerpohl JJ. Effect of lower sodium intake on health: Systematic review and meta-analyses. *BMJ*. 2013;346:f1326.

47. Marmot M, Brunner E. Alcohol and cardiovascular disease: The status of the U shaped curve. *BMJ*. 1991;303(6802):565-568.

48. Koning SH, Gansevoort RT, Mukamal KJ, et al. Alcohol consumption is inversely associated with the risk of developing chronic kidney disease. *Kidney Int*. 2015.

49. Lange T, Hansen JV. Direct and indirect effects in a survival context. *Epidemiology*. 2011;22(4):575-581.

50. Lange T, Rasmussen M, Thygesen LC. Assessing natural direct and indirect effects through multiple pathways. *Am J Epidemiol*. 2013;179(4):513-518.

Heart rate variability and its relation to chronic kidney disease: Results from the PREVEND Study

Christian HL Thio, Arie M van Roon, Joop D Lefrandt, Ron T Gansevoort, and Harold Snieder

Psychosom Med. 2018;80(3):307–16.

CHAPTER

3

ABSTRACT

Objective. In the general population, reduced heart rate variability (HRV) has been associated with cardiovascular disease. However, its relation to chronic kidney disease (CKD) is debated. We therefore investigated the relation between low HRV and renal outcomes.

Methods. In the population-based PREVEND Study, renal outcomes (CKD, eGFR, urinary albumin) were measured at baseline and three consecutive examinations. HRV measures (among which SDNN, standard deviation of normal-to-normal RR-intervals) were calculated from time-series of beat-to-beat pulse-wave recordings at baseline. The lowest (risk) quartile was compared to the upper three quartiles combined, in multivariable survival and linear mixed-effects analyses.

Results. In 4605 participants (49% males, age range 33-80, 0.6% blacks), we observed 341 new cases of CKD during a median follow-up duration of 7.4 years. Low SDNN was associated with higher incidence of CKD (crude HR: 1.66, 95%CI [1.30;2.12], p<0.001), but this association was no longer significant after adjustment for age, sex, and cardiovascular risk factors (adjusted HR: 1.13, 95%CI [0.86;1.48], p=0.40, similar for other HRV measures). No associations between SDNN and eGFR trajectories were found in the total sample. However, in a subgroup of participants with baseline CKD (N=939), we found a significant association of low SDNN (but not other HRV measures) with lower baseline eGFR, even after multivariable adjustment (adjusted $\beta_{\text{level difference}}$=-3.73 ml/min/1.73m^2, 95%CI [-6.70;-0.75], p=0.014), but not with steeper eGFR decline.

Conclusions. These results suggest that reduced HRV may be a complication of CKD rather than a causal factor.

LIST OF ABBREVIATIONS

CKD = chronic kidney disease
eGFR = estimated glomerular filtration rate
HF = high frequency power
HRV = heart rate variability
LF = low frequency power
PREVEND Study= Prevention of REnal and Vascular ENdstage Disease Study
rMSSD = root mean square of successive differences
SDNN = standard deviation of normal-to-normal RR-intervals
UAE = urinary albumin excretion

BACKGROUND

Chronic kidney disease (CKD) is a group of heterogeneous disorders characterized by kidney damage and impaired renal function, and is defined by an elevated urinary albumin excretion (UAE), a decreased glomerular filtration rate (GFR), or a combination of both.[1-3] The most important risk factors for CKD are diabetes and hypertension. However, it has been observed that CKD can also occur in the absence of these risk factors.[4,5] This suggests that other mechanisms may be involved in the development of CKD.

A potential causal mechanism involves imbalance of the autonomic nervous system, in which parasympathetic function is decreased relative to sympathetic function. Hypothetically, autonomic imbalance causes renal damage through changes in renal hemodynamics. In animal studies, stimulation of renal sympathetic afferents affected renal hemodynamics, while renal (sympathetic) denervation in these animals attenuated progression of kidney failure.[6-8] In humans, a non-invasive way of assessing autonomic function is by calculating heart rate variability (HRV), a measure of autonomic control over heart rate. It is the variation in duration between normal-to-normal (NN) RR-intervals.[9-12] Moderate-to-high HRV indicates healthy autonomic function, while low HRV reflects poor autonomic function, and has been associated with cardiovascular risk factors and adverse cardiovascular outcomes.[10,11,13-16] The relation between HRV and CKD has been explored in several small-scale studies. Participants with CKD were found to have lower HRV compared to those without CKD. Also, low HRV was associated with adverse outcomes during follow-up (i.e. progression to end-stage renal disease and mortality) in CKD patients, although results are inconsistent between studies.[17-23] The mechanisms underlying this association are still under investigation, but it is commonly believed that autonomic imbalance is a complication of renal damage[24].

However, in the Atherosclerosis Risk in Communities (ARIC)-cohort, a 20-108% higher incidence of CKD-related hospitalization and/or end-stage renal disease (ESRD) was observed in those with low HRV (first quartile) compared to those with normal-to-high HRV (upper three quartiles combined), even in participants with normal kidney function at baseline.[25] This suggests that autonomic imbalance may also play a role in the pathophysiology of CKD. To our knowledge, this finding has not yet been verified in other population-based longitudinal studies. If

autonomic imbalance is identified as a mechanism of renal damage, this may lead to improved risk prediction and novel therapeutic options.

In this study, our primary aim was therefore to investigate the association between HRV and new-onset CKD in a sample of the general population. Furthermore, we assessed whether low HRV was associated with baseline levels of eGFR and UAE and change in these parameters during follow-up.

METHODS

Study sample and design

We used data from the Prevention of REnal and Vascular ENdstage Disease (PREVEND) cohort study. Details of this study have been described elsewhere.26 In brief, 8592 individuals, sampled from the general population of Groningen, the Netherlands, completed an extensive examination between 1997- 1998. The second, third, fourth and fifth examination were completed in 2003, 2006, 2008 and 2012, respectively. For the present study, we refer to the second examination as 'baseline', as this was the first examination that included additional beat-to-beat blood pressure recordings that were used for calculation of HRV parameters. This examination was attended by 6894 participants, of which 2289 had missing HRV measures (due to either technical failure (N=397) or due to poor quality signal or excessive amount of artifacts in the recording (N=1892)), leaving 4605 participants for the present analyses. All participants gave written informed consent. The PREVEND Study was approved by the medical ethics committee of the University Medical Center Groningen, and conducted in accordance with the Helsinki Declaration guidelines.

Measurement

HRV measures

Details of the HRV measurement procedure in the PREVEND study have been described previously [27]. In brief, participants were measured in a supine position, in a quiet room kept at a constant temperature of 22°C. Participants were not allowed to talk or move during the procedure. Beat-to-beat heart rate was assessed by non-invasive 15-min pulse wave measurement using a Portapres® device (FMS Finapres Medical systems BV, Amsterdam, The Netherlands)[28] at baseline. From these 15-min measurements, we selected the last 4 to 5 minutes of stationary time-series of pulse wave data. Using CARSPAN v2.0 software [29], these time-

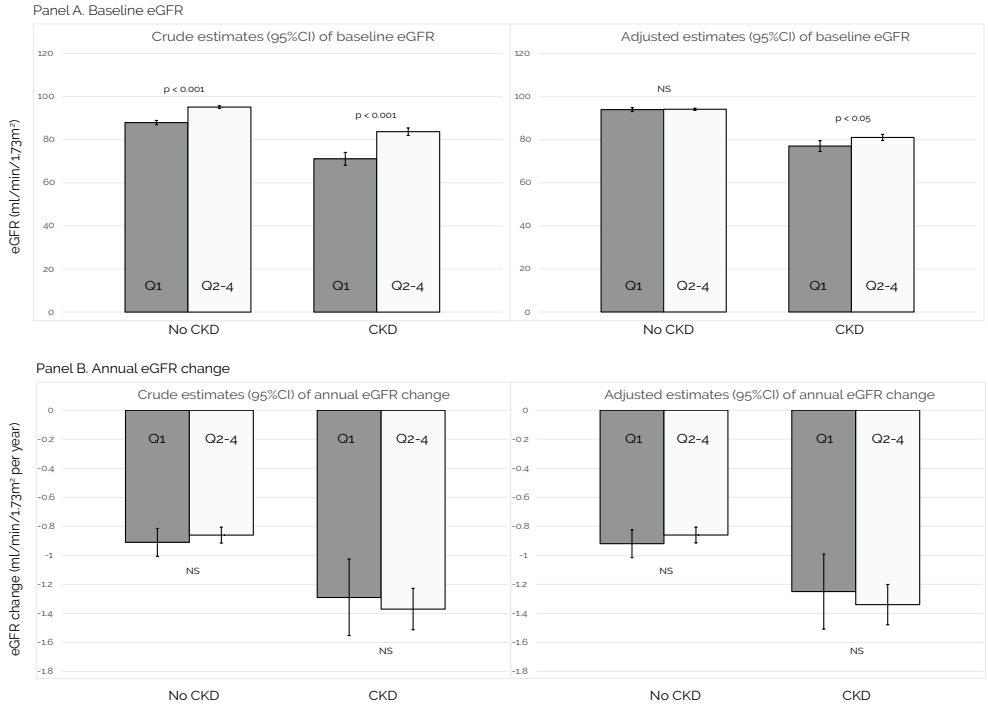
series were visually pre-processed to exclude cardiac arrhythmias, artefacts, electrical 'noise', or aberrant beats. Normal-to-normal RR-intervals from the beat-to-beat pulse wave signals were detected with an accuracy of 5ms (sampling frequency of 200 Hz). Artifacts were removed and the resulting gaps interpolated as described previously[30]. After pre-processing, HRV measures were calculated using the same CARSPAN software. HRV measures included standard deviation of normal-to-normal RR-intervals (SDNN) and root mean square of successive differences between normal-to-normal RR-intervals (rMSSD). To quantify cyclic changes in heart rate, we calculated high frequency (HF) and low frequency (LF) power (area under the power spectral density curve) by Fourier spectral analysis, and the ratio between LF/HF. LF power was defined as the total area between 0.04 and 0.15Hz, and HF power was defined as the total area between 0.15 and 0.40Hz[9-12]. HRV was categorized into low (lowest quartile, Q1) and moderate-to-high (upper three quartiles combined, Q2-4) to allow direct comparison to the work of Brotman et al[25].

Renal outcomes

Details of the assessment of eGFR and UAE have been described elsewhere[31]. In brief, participants collected two consecutive 24h-urine specimens at each screening round. The collected urine was stored cold (4°C) for a maximum of four days before handing it in. After this, urine specimens were stored at -20°C. Furthermore, fasting blood samples were obtained and stored at -80°C.

Measurement of serum creatinine was performed by an enzymatic method on a Roche Modular analyzer using reagents and calibrators from Roche (Roche Diagnostics, Mannheim, Germany), with intra- and interassay coefficients of variation of 0.9% and 2.9%, respectively. Serum cystatin C concentration was measured by a Gentian cystatin C Immunoassay (Gentian AS Moss, Norway) on a Modular analyzer (Roche Diagnostics). Cystatin C was calibrated directly using the standard supplied by the manufacturer (traceable to the International Federation of Clinical Chemistry Working Group for Standardization of Serum Cystatin C)[32]. The intra- and interassay coefficients of variation were <4.1% and <3.3%, respectively. Urinary albumin concentration (UAC) was measured by nephelometry with a lower threshold of detection of 2.3mg/L, and intra- and interassay coefficient of variation of 2.2% and 2.6%, respectively (Dade Behring Diagnostic, Marburg, Germany). UAC was multiplied by urine volume to obtain a value of UAE in mg/24h. The two 24h

Figure 1. Associations of SDNN (Q1 vs Q2-4) with baseline eGFR level (panel A) and annual change in eGFR (panel B).



SDNN: standard deviation of normal-to-normal RR intervals; eGFR: estimated glomerular filtration rate; CKD: chronic kidney disease; NS: non-significant (p > 0.05). Covariates were centered to obtain adjusted estimates. Due to centering, estimates may differ slightly from Table 4.

UAE values of each subject per examination were averaged. eGFR was calculated according to the 2012 CKD-EPI creatinine-cystatin C equation[33]. CKD was defined as an eGFR<60ml/min/1.73m², a UAE≥30mg/24h, or both, according to the 2011 revised KDIGO guidelines[2].

Covariates

Known cardiovascular risk factors were included as covariates and assessed at baseline. Body-mass index (BMI: weight/height²) and waist-hip circumference ratio (WHR) were calculated from anthropometrics. Mean inter-beat interval (IBI) was calculated from time-series of beat-to-beat heart rate data. Smoking status was defined as self-reported never, former, or current smoker (subdivided in <6 cigarettes, 6-20 cigarettes, and >20 cigarettes daily). History of cardiovascular disease (CVD) was assessed using questionnaires, and was defined as a history of any cardio- or cerebro-vascular events. Hypertension was defined as SBP≥140mmHg, DBP≥90mmHg, or self-reported or pharmacy-reported prescribed use of blood pressure-lowering drugs, including ACE-inhibitors,

angiotensin-II receptor antagonists, beta blocking agents, diuretics (ATC codes 2, 3, 7, 8, 9). Diabetes was defined as either a fasting glucose level of >7mmol/L, or self-reported or pharmacy-reported prescribed use of anti-diabetic drugs. Hypercholesterolemia was defined as a total cholesterol ≥6.21mmol/L, or self-reported use or pharmacy reported prescribed use of lipid-lowering drugs.

Statistical analysis

Statistical analyses were performed using SPSS version 22.0 (IBM corporation). Two-sided significance level was set at α=0.05.

Baseline characteristics

Baseline characteristics were compared between HRV categories using Student's t-tests, Mann-Whitney U-tests, and χ^2-tests where appropriate.

Association of HRV with CKD incidence

For this analysis participants with CKD (N=939) or unknown CKD status at baseline (N=269) were excluded. Participants were censored at death, loss to follow-up, withdrawal, or end of study. We used mid-point imputation to approximate time to event.[34] Mantel-Cox log-rank tests were performed to test for equality in hazard rates between low HRV and moderate-to-high HRV. In Cox-regression models, we adjusted for potential confounders by introducing blocks of covariates. Block 1 included age; block 2 in addition included sex, BMI, WHR, mean IBI, smoking, baseline eGFR, and baseline UAE; block 3 additionally included history of CVD, diabetes, hypertension, and hypercholesterolemia. All covariates were retained in the model; no criteria for covariate exclusion were applied.

Association of HRV with baseline levels and change in eGFR and UAE

To examine the association of baseline HRV with eGFR and UAE over time, we conducted multivariable linear mixed-effects (LME)analyses in the entire sample (N=4,605). eGFR and the natural logarithm of UAE were modelled as a function of time. Based on model fit criteria and likelihood ratio tests, we specified a base model with unstructured covariance structure, random intercept, and random slope for time.

HRV category (Q1 vs Q2-4) was added to the model to assess its association with baseline eGFR and UAE. A two-way interaction between HRV and time was introduced to assess the association of HRV with change in eGFR (ml/min/1.73m^2 per year) and

UAE (mg/24h per year). In multivariable models, we adjusted for incremental blocks of covariates as described above.

Sensitivity analyses

By design, participants with a moderately elevated urinary albumin concentration (>10mg/L) are overrepresented in the PREVEND study. To address this imbalance, we performed sensitivity analyses using statistical weights that were based on the selection probability. Also, we performed 40 imputations using the fully conditional specification method[35,36], by which we imputed missing HRV and covariate data. Additional analyses included definitions of new-onset CKD based on either impaired eGFR only (CKD_{eGFR}: eGFR<60ml/min/1.73m²) or elevated UAE only (CKD_{UAE}: UAE≥30mg/24h). Furthermore, we applied a stricter definition of the high risk group by assigning to it participants that were in Q1 of each of the three main HRV parameters, SDNN, rMSSD, and HF ("Composite low HRV", see Figure S1, Supplemental Digital Content 1). Finally, we conducted analyses on continuous measures of HRV. For these analyses all HRV parameters were transformed by their natural logarithm, which improved linearity of the associations.

RESULTS

Baseline characteristics

Baseline characteristics of the 4605 participants are presented in **Table 1**, stratified according to low vs moderate-to-high HRV (Q1 vs Q2-4), for SDNN, rMSSD, and HF (for LF, LF/HF-ratio see Table S1a, SDC 1). The medians (IQR) of the different HRV parameters are listed in **Table 2**. In univariable analyses, participants in Q1 of SDNN had lower eGFR, higher UAE at baseline, and were more likely to have CKD at baseline. Those with baseline CKD had mildly diminished eGFR (mean(sd)=81(22); eGFR<60 in 20%) and elevated UAE (mean[IQR]=43[24-89]; UAE≥30 in 70%; see **Table S2**, SDC 1). In Q1 of SDNN we observed a less favorable cardiovascular risk profile compared to Q2-4, i.e. higher prevalence of diabetes, hypertension, hypercholesterolemia, current smoking, and history of CVD. Similar results were found for other HRV measures.

In univariable comparisons between the 4605 included participants and the 2289 excluded participants of whom no valid HRV measurements were available, no relevant differences were observed in covariates or outcomes (data not shown).

Table 1. Baseline characteristics by heart rate variability categories (Q1 vs Q2-4) for the entire sample.

		SDNN			rMSSD			HF		
		Q1	Q2-4		Q1	Q2-4		Q1	Q2-4	
	Total	4.6-23 ms	23-262 ms	p	6.4-17 ms	17-377 ms	p	3.9-94 ms²	>94ms²	p
N	4605	1151	3454	n/a	1151	3454	n/a	1151	3454	n/a
Age, years	53 [45-63]	61 [53-70]	50 [43-59]	**<0.001***	60 [52-69]	51 [43-60]	**<0.001***	60 [53-69]	51 [43-59]	**<0.001***
Males, n	2270 (49%)	592 (51)%	1678 (49%)	0.094	527 (46%)	1808 (52%)	**<0.001***	519 (45%)	1816 (53%)	**<0.001***
Black race, n	28 (0.6%)	9 (0.8%)	19 (0.6%)	0.38	9 (0.8%)	19 (0.6%)	0.38	8 (0.7%)	20 (0.6%)	0.66
Height, cm	173 (9.5)	171 (9.6)	173 (9.4)	**<0.001***	172 (9.3)	173 (9.6)	**<0.001***	172 (9.4)	173 (9.5)	**<0.001***
BMI, kg/m²	26.8 (4.4)	28 (4.7)	26 (4.2)	**<0.001***	27 (4.5)	27 (4.3)	**<0.001***	27 (4.5)	27 (4.3)	**<0.001***
WHR	0.90 (0.085)	0.92 (0.081)	0.89 (0.085)	**<0.001***	0.92 (0.082)	0.90 (0.085)	**<0.001***	0.92 (0.082)	0.89 (0.084)	**<0.001***
Heart rate, beats/min	68 (10)	74 (11)	66 (8.9)	**<0.001***	75 (10)	66 (8.8)	**<0.001***	75 (10)	66 (9.0)	**<0.001***
Smoking				**<0.001***			0.28			0.23
Never, n	1315 (29%)	287 (25%)	1028 (30%)		311 (27%)	1004 (29%)		315 (28%)	1000 (29%)	
Former, n	1934 (43%)	474 (42%)	1460 (43%)		482 (43%)	1452 (43%)		478 (42%)	1456 (43%)	
Current, n	1298 (29%)	374 (33%)	924 (27%)		432 (30%)	956 (28%)		347 (30%)	951 (28%)	
SBP, mmHg	127 (19)	133 (19)	124 (18)	**<0.001***	133 (20)	124 (18)	**<0.001***	134 (19)	124 (18)	**<0.001***
DBP, mmHg	74 (9.1)	76 (8.9)	73 (9.0)	**<0.001***	77 (9.2)	72 (8.8)	**<0.001***	77 (9.0)	72 (8.8)	**<0.001***
Antihypertensive R_x, n	1019 (25%)	386 (36%)	633 (21%)	**<0.001***	335 (31%)	684 (23%)	**<0.001***	347 (32%)	672 (22%)	**<0.001***
Hypertension , n	1578 (38%)	582 (53%)	996 (33%)	**<0.001***	546 (50%)	1032 (34%)	**<0.001***	563 (52%)	1015 (33%)	**<0.001***
Fasting glucose, mmol/L	4.8 [4.4-5.3]	5.0 [4.5-5.6]	4.7 [4.4-5.2]	**<0.001***	5.0 [4.5-5.0]	4.7 [4.4-5.3]	**<0.001***	5.0 [4.5-5.5]	4.7 [4.4-5.3]	**<0.001***
Antidiabetic R_x, n	169 (4.2%)	89 (8.3%)	80 (2.7%)	**<0.001***	84 (7.9%)	85 (2.9%)	**<0.001***	86 (8.1%)	83 (2.8%)	**<0.001***
Diabetes Mellitus , n	299 (7.5%)	137 (13%)	162 (5.6%)	**<0.001***	126 (12%)	173 (5.9%)	**<0.001***	122 (12%)	177 (6.0%)	**<0.001***
History of CVD, *n*	302 (6.8%)	118 (11%)	184 (5.5%)	**<0.001***	89 (8.0%)	213 (6.4%)	0.059	100 (9.0%)	202 (6.0%)	**0.001***
Total cholesterol, mmol/L	5.5 (1.0)	5.6 (1.0)	5.4 (1.0)	**<0.001***	5.7 (1.1)	5.4 (1.0)	**<0.001***	5.7 (1.0)	5.4 (1.0)	**<0.001***
Lipid lowering R_x, n	465 (11%)	193 (18%)	272 (9.1%)	**<0.001***	158 (15%)	307 (10%)	**<0.001***	175 (17%)	290 (9.7%)	**<0.001***
Hypercholesterolemia, n	1453 (35%)	497 (45%)	956 (32%)	**<0.001***	473 (43%)	980 (32%)	**<0.001***	477 (44%)	976 (32%)	**<0.001***
Serum creatinine, mg/dL	0.82 (0.23)	0.84 (0.32)	0.82 (0.18)	0.11	0.85 (0.32)	0.81 (0.19)	**<0.001***	0.85 (0.32)	0.81 (0.19)	**<0.001***
Serum cystatin C, mg/L	0.91 (0.21)	0.99 (0.29)	0.88 (0.18)	**<0.001***	0.98 (0.28)	0.89 (0.18)	**<0.001***	0.98 (0.28)	0.89 (0.37)	**<0.001***
eGFR , ml/min/1.73m²	92 (17)*	84 (18)	94 (16)	**<0.001***	85 (18)	94 (16)	**<0.001***	85 (18)	94 (16)	**<0.001***
UAE, mg/24h	8.9 [6.2-17]	10 [6.8-22]	8.5 [6.0-15]	**<0.001***	10 [6.8-24]	8.5 [6.0-15]	**<0.001***	10 [6.8-24]	8.5 [6.0-15]	**<0.001***
Baseline CKD, n	939 (22%)	331 (30%)	608 (19%)	**<0.001***	336 (31%)	603 (19%)	**<0.001***	340 (31%)	599 (18%(	**<0.001***
Baseline $CKD_{eGFR<60}$, n	202 (4.7%)	97 (9.0%)	105 (3.3%)	**<0.001***	94 (8.8%)	972 (91%)	**<0.001***	100 (9.4%)	102 (3.2%)	**<0.001***
Baseline $CKD_{UAE\geq30}$, n	846 (18%)	283 (25%)	563 (16%)	**<0.001***	292 (26%)	554 (16%)	**<0.001***	294 (26%)	552 (16%)	**<0.001***

SDNN: standard deviation of all normal-normal RR-intervals; rMSSD: root mean square of successive differences of adjacent normal-to-normal RR-intervals; HF: high frequency power spectrum; BMI: body mass index; WHR: waist/hip ratio; SBP: systolic blood pressure; DBP: diastolic blood pressure; CVD: cardiovascular disease; Rx: medication use; eGFR: estimated glomerular filtration rate; UAE: urinary albumin excretion; CKD: chronic kidney disease, defined as estimated glomerular filtration rate (eGFR) < 60 mL/min/1.73m² or urinary albumin excretion (UAE)≥30 mg/24 hours. * indicates statistical significance (p<0.05)

Table 2. Distribution of HRV parameters.	
SDNN (ms)	31 [23-42]
rMSSD (ms)	24 [17-35]
HF (ms^2)	211 [94-454]
LF (ms^2)	242 [123-494]
LF/HF-ratio	1.2 [0.7-2.0]

HRV measures were non-normally distributed, hence data is presented as median [interquartile range]. SDNN: standard deviation of normal-to-normal RR-intervals; rMSSD: root mean square of successive differences; HF: high frequency power spectrum; LF: low frequency power spectrum.

Association of HRV with CKD incidence

We excluded those with CKD or unknown CKD status at baseline, leaving 3397 participants. Baseline characteristics for these 3397 participants are presented in Table S1b-c, SDC 1. Of these participants, 341 developed CKD during a median of 7.4 [IQR: 7.0–7.8] years of follow-up. At the earliest moment of identification, those with new-onset CKD had mildly diminished eGFR (mean[IQR]=79[59-94]; eGFR<60 in 20%) and elevated UAE (mean[IQR]=35[17-48], UAE≥30 in 72%, see Table S2, SDC 1). Event rates of CKD per HRV category are shown in **Table 3**.

Table 3. Chronic kidney disease incidence rates by heart rate variability categories (Q1 vs Q2-4).

	Total	SDNN		p	rMSSD		p	HF		p
		Q1	Q2-4		Q1	Q2-4		Q1	Q2-4	
N	3397	849	2548	n/a	849	2548	n/a	849	2548	n/a
Person-years, [IQR]	6.1 [4.6-7.3]	5.4 [2.1-7.3]	6.8 [3.1-7.4]	**<0.001**	5.9 [2.1-7.3]	6.8 [2.7-7.4]	**<0.001**	6.4 [2.3-7.4]	6.8 [3.0-7.4]	**<0.001**
New-onset CKD [a] /n (%)	341 (10%)	116 (14%)	225 (8.8%)	**<0.001**	107 (13%)	234 (9.2%)	**0.004**	109 (13%)	232 (9.1%)	**0.002**
New-onset CKD /1000 py	19.5	29.1	16.7	**<0.001**	25.9	17.5	**<0.001**	26.9	17.3	**<0.001**

Event rates by HRV category (low vs moderate-to-high HRV, Q1 vs Q2-4). SDNN: standard deviation of normal-to-normal RR-intervals; rMSSD: root mean square of successive differences of adjacent normal-normal RR-intervals; HF: high frequency power spectrum; IQR: interquartile range; CKD: chronic kidney disease; py: person-years. [a] Defined as estimated glomerular filtration rate (eGFR)<60 mL/min/1.73m^2 or urinary albumin excretion (UAE)≥30 mg/24 hours.

Incidence rate of CKD was significantly higher in those with low HRV (SDNN Q1 vs Q2-4: 29.1 v 16.7 cases per 1,000 person-years, Mantel-Cox log-rank test χ^2=23.9, df=1, p<0.001, similar for other HRV measures). The results of Cox-regression analyses are shown in **Table 4** (results for LF, LF/HF-ratio in Table S4a, SDC 1). Low HRV was associated with CKD incidence (SDNN Q1 vs Q2-4: unadjusted hazard ratio (HR)=1.66, 95%CI [1.30;2.12], similar for other HRV measures). After adjusting for confounders, this association was no longer significant (SDNN Q1 vs Q2-4: fully adjusted HR=1.13, 95%CI [0.86;1.48], similar for rMSSD, HF, and LF). Only for LF/HF-ratio a significant association was found, which remained after multivariable

adjustment (LF/HF-ratio Q1 vs Q2-4: fully adjusted HR=1.32 [1.01;1.71], p<0.043). Alternative definitions of new-onset CKD (incidence of either impaired eGFR, or of elevated UAE) yielded similar results (see **Table 4**).

Table 4. Association of heart rate variability measures (Q1 vs Q2-4) with incident chronic kidney disease.

CKD	SDNN Q1	p	rMSSD Q1	p	HF Q1	p
Unadjusted HR [95%CI]	1.66 [1.30 ; 2.12]	**<0.001***	1.51 [1.18 ; 1.93]	**0.001***	1.54 [1.20 ; 1.97]	**<0.001***
Adjusted HR [95%CI] [1]	1.02 [0.79 ; 1.32]	0.88	1.01 [0.78 ; 1.30]	0.97	0.99 [0.77 ; 1.28]	0.93
Adjusted HR [95%CI] [2]	1.10 [0.83 ; 1.45]	0.50	1.09 [0.82 ; 1.45]	0.57	1.04 [0.78 ; 1.37]	0.80
Fully adjusted HR [95%CI] [3]	1.13 [0.86 ; 1.48]	0.40	1.09 [0.82 ; 1.45]	0.55	1.02 [0.77 ; 1.35]	0.87
$CKD_{eGFR<60}$						
Unadjusted HR [95%CI]	2.44 [1.64 ; 3.63]	**<0.001***	1.92 [1.28 ; 2.88]	**0.002***	2.05 [1.37 ; 3.07]	**<0.001***
Adjusted HR [95%CI] [1]	1.05 [0.70 ; 1.59]	0.80	0.97 [0.64 ; 1.46]	0.88	0.97 [0.64 ; 1.46]	0.88
Adjusted HR [95%CI] [2]	0.90 [0.57 ; 1.42]	0.66	1.09 [0.68 ; 1.75]	0.71	0.83 [0.52 ; 1.32]	0.83
Fully adjusted HR [95%CI] [3]	0.93 [0.59 ; 1.46]	0.76	1.16 [0.72 ; 1.85]	0.54	0.89 [0.56 ; 1.41]	0.61
$CKD_{UAE\geq30}$						
Unadjusted HR [95%CI]	1.46 [1.09 ; 1.96]	**0.011***	1.43 [1.07 ; 1.92]	**0.016***	1.39 [1.04 ; 1.87]	**0.028***
Adjusted HR [95%CI] [1]	1.04 [0.76 ; 1.41]	0.82	1.07 [0.79 ; 1.45]	0.64	1.01 [0.75 ; 1.38]	0.93
Adjusted HR [95%CI] [2]	1.15 [0.83 ; 1.60]	0.40	1.23 [0.87 ; 1.73]	0.24	1.12 [0.80 ; 1.57]	0.51
Fully adjusted HR [95%CI] [3]	1.17 [0.84 ; 1.62]	0.35	1.22 [0.87 ; 1.71]	0.25	1.10 [0.79 ; 1.54]	0.56

Estimates of hazard ratios after multivariable Cox regression analysis. Reference group is moderate-to-high HRV (Q2-4). HR: hazard ratio; SDNN: standard deviation of normal-to-normal RR-intervals; rMSSD: root mean square of successive differences of adjacent normal-normal RR-intervals; HF: high frequency power spectrum; CI: confidence interval. * indicates statistical significance (p<0.05)

[1] Adjusted for age
[2] Adjusted for sex, BMI, WHR, mean IBI, smoking status, baseline eGFR, baseline UAE, in addition to above
[3] Adjusted for history of cardiovascular disease, diabetes, hypertension, and hypercholesterolemia, in addition to above.

Sensitivity analyses in imputed datasets (in which we imputed missing values of HRV and covariates), and analyses with sampling weights (to account for sampling imbalance), did not substantially change results for SDNN, rMSSD, HF, and LF (see Supplementary Tables S4b-d). However, the multivariable-adjusted HR for LF/HF-ratio was no longer significant in these analyses (LF/HF-ratio Q1 vs Q2-4: fully adjusted HR=1.19, 95%CI [0.79;1.79], in imputed datasets, similar for weighted analysis). Furthermore, a more stringent definition of the high risk group ("Composite low HRV", participants in Q1 of each of the main HRV parameters, SDNN, rMSSD, and HF, see Supplementary Table 4a) yielded similar results.

Association of HRV with baseline levels and change in eGFR and UAE

In **Table 5**, the results of LME analyses are shown for all 4605 participants (for LF and LF/HF-ratio, see Table S5a, SDC 1). Those with low HRV had significantly

Table 5. Differences between low (Q1) and moderate-to-high heart rate variability (Q2-4) measures for baseline levels and rate of decline of eGFR.

	SDNN Q1					
	Total (N=4605)	**p**	**No CKD (N=3397)**	**p**	**CKD (N=939)**	**p**
Baseline eGFR-level difference [a] (ml/min/1.73m²)						
Unadjusted β [95%CI]	-9.36 [-10.6 ; -8.08]	**<0.001***	-7.36 [-8.56 ; -6.17]	**<0.001***	-12.3 [-15.8 ; -8.74]	**<0.001***
Adjusted β [95%CI] [1]	-0.94 [-1.97 ; 0.092]	0.074	-0.60 [-1.59 ; 0.40]	0.24	-3.52 [-6.39 ; -0.66]	**0.016***
Adjusted β [95%CI] [2]	-0.81 [-1.90 ; 0.29]	0.15	-0.43 [-1.48 ; 0.63]	0.43	-4.02 [-7.05 ; -0.98]	**0.010***
Fully adjusted β [95%CI] [3]	-0.59 [-1.66 ; 0.48]	0.28	-0.42 [-1.48 ; 0.63]	0.43	-3.73 [-6.70 ; -0.75]	**0.014***
eGFR-slope difference [b] (ml/min/1.73m2 per year)						
Unadjusted β_{slope} [95%CI]	-0.068 [-0.18 ; 0.039]	0.21	-0.048 [-0.16 ; 0.063]	0.40	0.080 [-0.22 ; 0.38]	0.60
Adjusted β_{slope} [95%CI] [1]	-0.076 [-0.18 ; 0.031]	0.16	-0.061 [-0.17 ; 0.050]	0.28	0.075 [-0.22 ; 0.37]	0.62
Adjusted β_{slope} [95%CI] [2]	-0.072 [-0.18 ; 0.034]	0.18	-0.058 [-0.17 ; 0.053]	0.30	0.078 [-0.22 ; 0.37]	0.60
Fully adjusted β_{slope} [95%CI] [3]	-0.077 [-0.18 ; 0.029]	0.16	-0.059 [-0.17 ; 0.052]	0.30	0.086 [-0.21 ; 0.38]	0.57
	rMSSD Q1					
	Total (N=4605)	**p**	**No CKD (N=3397)**	**p**	**CKD (N=939)**	**p**
Baseline eGFR-level difference [a] (ml/min/1.73m²)						
Unadjusted β [95%CI]	-8.11 [-9.40 ; -6.82]	**<0.001***	-6.26 [-7.46 ; -5.05]	**<0.001***	-7.64 [-11.3 ; -3.98]	**<0.001***
Adjusted β [95%CI] [1]	-0.70 [-1.72 ; 0.32]	0.18	-0.51 [-1.48 ; 0.47]	0.31	-0.98 [-3.83 ; 1.87]	0.50
Adjusted β [95%CI] [2]	-0.90 [-2.02 ; 0.22]	0.11	-0.79 [-1.87 ; 0.29]	0.15	-1.42 [-4.56 ; 1.71]	0.37
Fully adjusted β [95%CI] [3]	-0.68 [-1.77 ; 0.42]	0.23	-0.83 [-1.91 ; 0.25]	0.13	-1.37 [-4.43 ; 1.69]	0.38
eGFR-slope difference [b] (ml/min/1.73m² per year)						
Unadjusted β_{slope} [95%CI]	-0.064 [-0.17 ; 0.043]	0.24	-0.055 [-0.17 ; 0.056]	0.33	0.22 [-0.080 ; 0.51]	0.15
Adjusted β_{slope} [95%CI] [1]	-0.068 [-0.17 ; 0.038]	0.21	-0.062 [-0.17 ; 0.048]	0.27	0.22 [-0.075 ; 0.51]	0.14
Adjusted β_{slope} [95%CI] [2]	-0.062 [-0.17 ; 0.044]	0.25	-0.059 [-0.17 ; 0.051]	0.29	0.22 [-0.075 ; 0.51]	0.15
Fully adjusted β_{slope} [95%CI] [3]	-0.064 [-0.17 ; 0.042]	0.24	-0.059 [-0.17 ; 0.051]	0.29	0.22 [-0.071 ; 0.51]	0.14
	HF Q1					
	Total (N=4605)	**p**	**No CKD (N=3397)**	**p**	**CKD (N=939)**	**p**
Baseline eGFR-level difference [a] (ml/min/1.73m²)						
Unadjusted β [95%CI]	-8.89 [-10.2 ; -7.60]	**<0.001***	-6.97 [-8.17 ; -5.77]	**<0.001***	-8.94 [-12.6 ; -5.29]	**<0.001***
Adjusted β [95%CI] [1]	-0.94 [-1.97 ; 0.085]	0.072	-0.66 [-1.64 ; 0.32]	0.19	-1.52 [-4.38 ; 1.35]	0.30
Adjusted β [95%CI] [2]	-1.11 [-2.22 ; 0.0022]	0.050	-0.82 [-1.88 ; 0.24]	0.13	-1.88 [-4.96 ; 1.20]	0.23
Fully adjusted β [95%CI] [3]	-0.76 [-1.84 ; 0.32]	0.17	-0.79 [-1.85 ; 0.27]	0.14	1.62 [-4.62 ; 1.39]	0.17
eGFR-slope difference [b] (ml/min/1.73m² per year)						
Unadjusted β_{slope} [95%CI]	-0.087 [-0.20 ; 0.021]	0.12	-0.065 [-0.18 ; 0.046]	0.25	0.21 [-0.093 ; 0.50]	0.18
Adjusted β_{slope} [95%CI] [1]	-0.090 [-0.20 ; 0.017]	0.10	-0.077 [-0.19 ; 0.034]	0.17	0.21 [-0.087 ; 0.50]	0.17
Adjusted β_{slope} [95%CI] [2]	-0.082 [-0.19 ; 0.025]	0.13	-0.075 [-0.19 ; 0.036]	0.18	0.21 [-0.089 ; 0.50]	0.17
Fully adjusted β_{slope} [95%CI] [3]	-0.087 [-0.19 ; 0.020]	0.11	-0.076 [-0.19 ; 0.035]	0.18	0.21 [-0.087 ; 0.50]	0.17

Estimates of the association between low HRV and eGFR in the total PREVEND population, and stratified for CKD at baseline, from multivariable linear mixed effects analysis. Reference group is moderate-to-high HRV (Q2-4), [a] eGFR-level: difference in baseline levels of eGFR, expressed in ml/min/1.73m², compared to reference, [b] eGFR-slope: difference in change in eGFR over time, in ml/min/1.73m² per year, compared to reference. HRV: heart rate variability; eGFR: estimated glomerular filtration rate; SDNN: standard deviation of normal-to-normal RR-intervals; rMSSD: root mean square of successive differences of adjacent normal-normal RR-intervals; HF: high frequency power spectrum; CI: confidence interval.

[1] Adjusted for age
[2] Adjusted for sex, BMI, WHR, mean IBI, smoking status, baseline UAE, in addition to above
[3] Adjusted for history of cardiovascular disease, diabetes, hypertension, hypercholesterolemia, (and baseline chronic kidney disease status in the total cohort) in addition to above.
* indicates statistical significance (p<0.05)

lower baseline levels of eGFR in the total sample (SDNN Q1 vs. Q2-4, unadjusted $\beta_{level\ difference}$=-9.36 ml/min/1.73m², 95%CI [-10.6;-8.08], p<0.001, similar for other HRV measures). However, after multivariable adjustment, the association of

low HRV with baseline eGFR was no longer significant (SDNN Q1 vs. Q2-4, fully adjusted $\beta_{\text{level difference}}$=-0.59 ml/min/1.73m², 95%CI [-1.66;0.48], p=0.28, similar for other HRV measures). During follow-up there was no significant difference in rate of decline of eGFR between HRV categories (SDNN Q1 vs Q2-4, fully adjusted $\beta_{\text{slope difference}}$=-0.077 ml/min/1.73m² per year, 95%CI [-0.18;0.029], p=0.16, similar for other HRV measures). Similarly, we found no significant association of HRV measures with UAE levels or increase (see Table S6a-b, SDC 1).

Next, we tested for a modifying effect of baseline CKD status on both level and slope by introducing their interaction terms (CKD*HRV*time; CKD*HRV; and CKD*time, in addition to their main effects) to the model. Addition of the interaction term resulted in a significant increase in log-likelihood (χ^2=64.5, Δdf=3, $p_{\text{interaction}}$<0.001 for SDNN, similar for other HRV measures), suggesting a modifying effect of baseline CKD status on the association between HRV and eGFR. Therefore, we stratified for baseline CKD status. For participants with CKD at baseline, low SDNN was associated with lower baseline eGFR. This cross-sectional association between SDNN and baseline eGFR remained after multivariable adjustment (SDNN Q1 vs Q2-4, fully adjusted $\beta_{\text{level difference}}$=-3.73 ml/min/1.73m², 95%CI [-6.70;-0.75], p=0.014). Other HRV measures did not show an association with lower baseline eGFR in this subgroup. There were no significant associations between low HRV measures and rate of renal function decline during follow-up (SDNN Q1 vs Q2-4, fully adjusted $\beta_{\text{slope difference}}$=0.086 ml/min/1.73m² per year, 95%CI [-0.21;0.38], p=0.57, similar for other HRV measures). In **Figure 1**, we show crude and adjusted estimates of baseline eGFR level (panel A) and annual eGFR change (panel B), by SDNN category and strata according to baseline CKD status.

Sensitivity analyses in imputed datasets (see Supplementary Tables S5b-c, S6c-d) yielded similar results. Application of a stricter definition of low HRV confirmed the significant result for SDNN (see Supplementary Tables S5a, S5c). Correlations (crude and age-adjusted) of HRV measures with kidney function outcomes reflected the results of our main analyses: 1) higher HRV correlated with higher baseline eGFR, but no longer after adjustment for age and 2) HRV showed no relevant correlations with eGFR slope (see **Table 6**). Results of Cox regression of continuous HRV measures supported our conclusions for the main outcome, CKD incidence. However, the association of continuous HRV with baseline levels of eGFR in CKD patients was not significant in these sensitivity analyses (see Table S7-8, SDC 1).

Table 6. Correlations between HRV parameters and kidney function outcomes

	eGFR		eGFR slope^	
	Crude	**Age-adjusted**	**Crude**	**Age-adjusted**
lnSDNN	0.276***	0.020	0.002	0.002
lnrMSSD	0.223***	-0.002	0.001	0.001
lnHF	0.254***	0.002	0.002	0.002
lnLF	0.310***	0.040**	0.003*	0.003*
lnLF/HF-ratio	0.044**	0.042**	0.001	0.000

Pearson's r and partial (age-adjusted) correlations between kidney function (eGFR and eGFR decline) and continuous, natural log-transformed HRV parameters in the total sample. * p<0.05, ** p<0.01, *** p<0.001

^ correlations for eGFR slope are standardized β's from linear mixed effects models.

All Supplementary material can be accessed using the following link www.links.lww.com/PSYMED/A436.

DISCUSSION

In this population-based, longitudinal cohort study, we examined the relation between HRVand renal outcomes. We observed an association between low HRV and higher incidence of CKD, which did not remain significant after adjustment for known CKD risk factors such as age, diabetes mellitus, and hypertension. The association between HRV and CKD risk could for a substantial part be explained by older age of those with lower HRV. An analysis of renal function over time in the total sample revealed no evidence for steeper decline in eGFR or increase in UAE in those with low HRV. In a subgroup of participants with CKD at baseline, for SDNN and a stricter definition of low HRV, we found a significant association with lower levels of baseline eGFR, which remained after adjustment for confounders, but no association with change in eGFR. For the other HRV measures (rMSSD, HF, LF, and LF/HF-ratio), we did not find significant associations with either baseline levels of eGFR or decline in eGFR during follow-up in this subgroup. These results suggest that low HRV does not contribute to CKD or to renal function decline. However, we observed that low HRV was associated with lower renal function in those that already have CKD. This implies another relation, i.e. CKD resulting in (or at least coinciding with) reduced HRV.

To our knowledge, the only comparable population-based study of HRV and its association with renal outcomes to date was conducted by Brotman et al[25]. In a sample of 13,241 adults of the ARIC cohort they observed that low HRV preceded

CKD-related hospitalization and ESRD. In our study, we could not corroborate these findings. Several differences may explain the inconsistent results. First, the endpoints and available measurements used are different: our endpoint was new-onset CKD (based on repeated measurements of serum creatinine, serum cystatin C, and UAE at each subsequent examination), whereas in ARIC, the endpoints were CKD hospitalization and ESRD. The endpoints used in ARIC imply more advanced renal disease, and are therefore a less suitable measure of de novo, likely mild, disease. Furthermore, due to the lack of baseline albumin measurements in their study, Brotman et al. could not exclude reverse causality, i.e. renal damage leading to low HRV. Second, there is a marked difference in study sample. The ARIC sample consisted of ~25% blacks, which accounted for ~50% of incident cases. This may have limited the comparability of their results to the PREVEND study, which consisted of only 0.6% blacks. A recent meta-analysis established that blacks, compared to whites, have on average higher resting values of HRV.[37] This is counter-intuitive, as black race has been associated with a higher cardiovascular risk profile[38] and risk of ESRD[39]. The ethnic differences suggest as yet unknown race-specific disease mechanisms, and stratified analyses may be warranted. Unfortunately, Brotman et al. did not explicitly adjust for race, or report race-stratified analyses. Therefore, it is unclear whether their findings also pertain to whites separately within ARIC.

Hypertension, diabetes, and cardiovascular disorders are possibly related to HRV in a bidirectional manner[13,40]. Therefore, the inclusion of these covariates in the statistical models may have led to underestimation of the effect of HRV. However, this is unlikely to have affected conclusions with regards to our main outcome, as inclusion of age almost completely explained the association between low HRV and incident CKD.

In CKD patients, we found low SDNN, and a stricter definition of low HRV, to be independently associated with lower baseline levels of eGFR, but not with steeper decline in eGFR in this subgroup. To our knowledge, the largest prospective study of HRV and disease outcomes in participants with CKD was performed by Drawz et al.[21] In 3,245 renal patients in the Chronic Renal Insufficiency Cohort (CRIC), HRV (calculated from 10s ECGs) was not independently associated with either ESRD or 50% decline in eGFR. Although we could not assess incidence of ESRD due to low numbers in our cohort, our finding that low HRV was not associated with steeper

eGFR decline is consistent with these results. In contrast, Chandra et al. did find a significant association of 24h LF/HF-ratio with incident ESRD in CKD patients.[20] However, this study was relatively small (N=305) and was a prognostic study on incidence of ESRD, rather than an etiological one, thus did not formally correct for potential confounders.[41]

In our sample of the general population, reduced HRV did not precede CKD. In contrast, we did observe an association of low SDNN, and of a stricter definition of low HRV, with low eGFR in participants that already had CKD, implying that reduced HRV is preceded by CKD. If there is any causal relationship between the two, it is more likely to be in a reversed direction (i.e. CKD causing reduced HRV). Salman recently reviewed several proposed mechanisms through which CKD could lead to increased sympathetic tone and/or decreased parasympathetic tone. Among others, these include: impaired reflex control of autonomic activity, activation of the renin-angiotensin-aldosterone system, activation of renal afferents, and mental stress in CKD[24]. Of noted interest is the potential role of social and psychological factors in the relation between CKD and HRV: e.g. mental stressors are proposed to contribute to the CKD risk factors, hypertension and diabetes, through alterations in autonomic nervous system activity and the neuro-endocrine system[42]. However, the pathophysiology underlying this relation is incompletely understood. Future work may include further characterization of these proposed mechanisms, in studies with repeated measures of autonomic and renal function as well as psychological and behavioral measures in race-stratified high-risk populations.

Major strengths of this study include the availability of serially measured creatinine and cystatin C based eGFR and 24h UAE values, which are considered to be the best parameters to define CKD, during considerable duration of follow-up. We examined multiple measures of HRV, calculated from time-series of highly standardized beat-to-beat recordings. To our knowledge, this is only the second study in the general population to examine the association of HRV with incidence of CKD, and the first to assess its effect on change in eGFR and UAE. This study is therefore an important contribution to the literature.

There were several limitations. First, HRV was calculated from time-series of pulse wave recordings. In individuals at rest, pulse rate variability is considered an accurate estimate of heart rate variability.[43] However, due to the lack of ECG

data we could not definitively exclude cardiac arrhythmias. Second, because follow-up HRV measurements were not available, we were unable to examine the association of HRV changes over time with renal disease, or vice versa. Third, HRV was missing in ~33% of participants. In an effort to minimize any bias introduced by the missingness, we conducted sensitivity analyses in multiple imputed datasets, the results of which did not change our conclusions. Although the missingness is likely random and non-problematic (e.g. due to technical failure, subject movement leading to artefacts in the recording) we cannot definitively rule out that in some cases, missing or invalid recordings may have been caused by non-random, unobserved mechanisms (e.g. cardiac arrhythmias). Fourth, estimates of GFR are less accurate in the higher range (>60 mL/min/1.73m^2). We therefore used the CKD-EPI equation for both creatinine and cystatin C, currently the best option for population-based studies.[33] Fifth, we lacked specific information on -blocking agents. This class of antihypertensive medication potentially affects both HRV and kidney function, and may therefore have caused unobserved confounding. However, we estimate -blocker user baseline prevalence to be low in this relatively healthy sample of the general population, and do not expect our conclusions to be substantially affected.

These results challenge the notion that reduced HRV represents a causal factor in CKD. Rather, they suggest that reduced HRV may be a complication of CKD.

REFERENCES

1. Chronic Kidney Disease Prognosis Consortium, Matsushita K, van der Velde M, et al. Association of estimated glomerular filtration rate and albuminuria with all-cause and cardiovascular mortality in general population cohorts: A collaborative meta-analysis. *Lancet*. 2010;375(9731):2073-2081.

2. Levey AS, de Jong PE, Coresh J, et al. The definition, classification, and prognosis of chronic kidney disease: A KDIGO controversies conference report. *Kidney Int*. 2011;80(1):17-28.

3. Levey AS, Coresh J. Chronic kidney disease. *Lancet*. 2012;379(9811):165-180.

4. 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):180-188.

5. Ozyilmaz A, de Jong PE, Gansevoort RT. Screening for chronic kidney disease can be of help to prevent atherosclerotic end-organ damage. *Nephrol Dial Transplant*. 2012;27(11):4046-4052.

6. DiBona GF. Neural control of the kidney: Past, present, and future. *Hypertension*. 2003;41(3 Pt 2):621-624.

7. Joles JA, Koomans HA. Causes and consequences of increased sympathetic activity in renal disease. *Hypertension*. 2004;43(4):699-706.

8. DiBona GF. Physiology in perspective: The wisdom of the body. neural control of the kidney. *Am J Physiol Regul Integr Comp Physiol*. 2005;289(3):R633-41.

9. Akselrod S, Gordon D, Ubel FA, Shannon DC, Berger AC, Cohen RJ. Power spectrum analysis of heart rate fluctuation: A quantitative probe of beat-to-beat cardiovascular control. *Science*. 1981;213(4504):220-222.

10. Heart rate variability: Standards of measurement, physiological interpretation and clinical use. task force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. *Circulation*. 1996;93(5):1043-1065.

11. Routledge HC, Chowdhary S, Townend JN. Heart rate variability--a therapeutic target? *J Clin Pharm Ther*. 2002;27(2):85-92.

12. Kleiger RE, Stein PK, Bigger JT,Jr. Heart rate variability: Measurement and clinical utility. *Ann Noninvasive Electrocardiol*. 2005;10(1):88-101.

13. Thayer JF, Lane RD. The role of vagal function in the risk for cardiovascular disease and mortality. *Biol Psychol*. 2007;74(2):224-242.

14. Thayer J, Hansen A, Johnsen B. Noninvasive assessment of autonomic influences on the heart: Impedance cardiography and heart rate variability. In: Luecken L, Gallo L, eds. *Handbook of physiological research methods in health psychology.* ; 2008:183-211.

15. Thayer JF, Yamamoto SS, Brosschot JF. The relationship of autonomic imbalance, heart rate variability and cardiovascular disease risk factors. *Int J Cardiol*. 2010;141(2):122-131.

16. Hillebrand S, Gast KB, de Mutsert R, et al. Heart rate variability and first cardiovascular event in populations without known cardiovascular disease: Meta-analysis and dose–response meta-regression. *Europace*. 2013;15(5):742-749.

17. Tory K, Suveges Z, Horvath E, et al. Autonomic dysfunction in uremia assessed by heart rate variability. *Pediatr Nephrol*. 2003;18(11):1167-1171.

18. Fukuta H, Hayano J, Ishihara S, et al. Prognostic value of heart rate variability in patients with end-stage renal disease on chronic haemodialysis. *Nephrol Dial Transplant*. 2003;18(2):318-325.

19. Burger AJ, D'Elia JA, Weinrauch LA, Lerman I, Gaur A. Marked abnormalities in heart rate variability are associated with progressive deterioration of renal function in type I diabetic patients with overt nephropathy. *Int J Cardiol*. 2002;86(2-3):281-287.

20. Chandra P, Sands RL, Gillespie BW, et al. Predictors of heart rate variability and its prognostic significance in chronic kidney disease. *Nephrol Dial Transplant*. 2012;27(2):700-709.

21. Drawz PE, Babineau DC, Brecklin C, et al. Heart rate variability is a predictor of mortality in chronic kidney disease: A report from the CRIC study. *Am J Nephrol*. 2013;38(6):517-528.

22. Chandra P, Sands RL, Gillespie BW, et al. Relationship between heart rate variability and pulse wave velocity and their association with patient outcomes in chronic kidney disease. *Clin Nephrol*. 2014;81(1):9-19.

23. Oikawa K, Ishihara R, Maeda T, et al. Prognostic value of heart rate variability in patients with renal failure on hemodialysis. *Int J Cardiol*. 2009;131(3):370-377.

24. Salman IM. Cardiovascular autonomic dysfunction in chronic kidney disease: A comprehensive review. *Curr Hypertens Rep*. 2015;17(8):59-015-0571-z.

25. Brotman DJ, Bash LD, Qayyum R, et al. Heart rate variability predicts ESRD and CKD-related hospitalization. *J Am Soc Nephrol*. 2010;21(9):1560-1570.

26. Pinto-Sietsma SJ, Janssen WM, Hillege HL, Navis G, De Zeeuw D, De Jong PE. Urinary albumin excretion is associated with renal functional abnormalities in a nondiabetic population. *J Am Soc Nephrol*. 2000;11(10):1882-1888.

27. Munoz ML, van Roon A, Riese H, et al. Validity of (ultra-)short recordings for heart rate variability measurements. *PLoS ONE*. 2015;10(9):e0138921.

28. Wesseling KH. Finapres: Continuous noninvasive finger arterial pressure based on the method of peňáz. In: Meyer-Sabellek W, Anlauf M, Gotzen R, Steinfeld L, eds. *Blood pressure measurement.* 1st ed. Darmstadt: Steinkopf Verlag; 1990:161-172.

29. Mulder, L. J. M.,, Roon, A. M. van,, Schweizer,D.A.,. *CARSPAN : Cardiovascular data analysis environment : User's manual.* Groningen: Iec ProGAMMA; 1995.

30. Greaves-Lord K, Tulen J, Dietrich A, et al. Reduced autonomic flexibility as a predictor for future anxiety in girls from the general population: The TRAILS study. *Psychiatry Res*. 2010;179(2):187-193.

31. Koning SH, Gansevoort RT, Mukamal KJ, et al. Alcohol consumption is inversely associated with the risk of developing chronic kidney disease. *Kidney Int*. 2015.

32. Grubb A, Blirup-Jensen S, Lindstrom V, et al. First certified reference material for cystatin C in human serum ERM-DA471/IFCC. *Clin Chem Lab Med*. 2010;48(11):1619-1621.

33. Inker LA, Schmid CH, Tighiouart H, et al. Estimating glomerular filtration rate from serum creatinine and cystatin C. *N Engl J Med*. 2012;367(1):20-29.

34. Law CG, Brookmeyer R. Effects of mid-point imputation on the analysis of doubly censored data. *Stat Med*. 1992;11(12):1569-1578.

35. Montez-Rath ME, Winkelmayer WC, Desai M. Addressing missing data in clinical studies of kidney diseases. *Clin J Am Soc Nephrol*. 2014;9(7):1328-1335.

36. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Methods Med Res*. 2007;16(3):219-242.

37. Hill LK, Hu DD, Koenig J, et al. Ethnic differences in resting heart rate variability: A systematic review and meta-analysis. *Psychosom Med*. 2015;77(1):16-25.

38. Mensah GA, Mokdad AH, Ford ES, Greenlund KJ, Croft JB. State of disparities in cardiovascular health in the united states. *Circulation*. 2005;111(10):1233-1241.

39. Saran R, Li Y, Robinson B, et al. US renal data system 2015 annual data report: Epidemiology of kidney disease in the United States. *Am J Kidney Dis*. 2016;67(3 Suppl 1):Svii, S1-305.

40. Acharya UR, Joseph KP, Kannathal N, Lim CM, Suri JS. Heart rate variability: A review. *Medical and biological engineering and computing*. 2006;44(12):1031-1051.

41. Tripepi G, Jager KJ, Dekker FW, Zoccali C. Testing for causality and prognosis: Etiological and prognostic models. *Kidney Int*. 2008;74(12):1512-1515.

42. Bruce MA, Beech BM, Sims M, et al. Social environmental stressors, psychological factors, and kidney disease. *J Investig Med*. 2009;57(4):583-589.

43. Schäfer A, Vagedes J. How accurate is pulse rate variability as an estimate of heart rate variability?: A review on studies comparing photoplethysmographic technology with an electrocardiogram. *Int J Cardiol*. 2013;166(1):15-29.

PART II

Genetics of kidney function and
the translation to clinical and research practice

Familial aggregation of chronic kidney disease and heritability of renal biomarkers in the general population: The Lifelines Cohort Study

Jia Zhang, Chris HL Thio, Ron T Gansevoort, Harold Snieder

Under review

CHAPTER

4

ABSTRACT

Introduction. Chronic kidney disease (CKD) is a major burden on patients and health resources, with a heritable component. We aimed to quantify familial aggregation of CKD in the general population, and assess the extent to which kidney traits can be explained by genetic or environmental factors.

Methods. This cross-sectional family study used baseline data from the Lifelines Cohort study, a sample of the general population of the Northern Netherlands with a unique three-generation design. CKD was defined by estimated glomerular filtration rate (eGFR) <60mL/min/1.73m^2 (1862 cases) and/or urinary albumin excretion (UAE) ≥30 mg/24hr (4127 cases). eGFR was calculated by CKD-EPI equation for serum creatinine (N=155,911). UAE was determined in 24h urine collections in a subsample (N=59,943). To quantify familial aggregation of CKD we calculated the recurrence risk ratio (RRR) with Cox proportional hazards models. Heritability of continuous kidney-related traits was estimated using linear mixed models. All models were adjusted for age, sex, and known renal risk factors.

Results. RRR of CKD in case of an affected first-degree relative was 3.05 (95%CI: 2.27-4.11), i.e. risk of CKD was 3.05 times higher compared to risk in the general population. In case of a spouse with CKD, the RRR was 1.61 (95% CI: 1.29-2.11), indicative of shared environmental factors and/or assortative mating. We report heritability estimates of eGFR (44%), UAE (20%), serum urea (31%), creatinine (37%) and uric acid (48%), and serum electrolytes (range 22%-28%).

Conclusions. In this large population-based family study, a positive family history was strongly associated with increased risk of CKD. We observed moderate to high heritability of renal traits and related biomarkers. These results indicate an important role of genetic factors in CKD risk and may inform preventive policies.

Keywords. familial aggregation; heritability; chronic kidney disease (CKD); kidney function; albuminuria

INTRODUCTION

Chronic kidney disease (CKD) is recognized as a global public health problem[1], with prevalence ranging between 3.3% and 17.3% in adult European populations[2]. Chronic kidney disease is defined by reduced estimated glomerular filtration rate (eGFR) and/or increased albuminuria, and is associated with an increased risk of cardiovascular disease and progression to end stage kidney disease (ESKD)[3-6].

Established risk factors for CKD, such as hypertension and diabetes, explain 50-70% of cases, and are the main targets of current risk prediction models for CKD[7]. Familial clustering of CKD and kidney related markers suggests that genetic factors or shared environmental factors are also important in the pathogenesis of this disease[8-12]. Support for a genetic component to CKD comes from recent genome-wide association studies (GWASs) on eGFR[13] and albuminuria[14], in which a large number of genetic loci have been reported. Despite the evidence for this genetic influence, the magnitude of the familial contribution to CKD susceptibility in the general population is poorly known.

One option of assessing the magnitude of the familial contribution to CKD is to examine its aggregation in families. Most familial aggregation studies on CKD focused on its later stages, i.e. end-stage kidney disease (ESKD) using medical records and registry data[8,9,12,15]. Focusing on early-stage CKD rather than ESKD may lead to more accurate estimates of familial recurrence risk of CKD, and may have added value for primary and secondary prevention strategies.

In addition to assessing familial recurrence risk, one can estimate the heritability of disease traits. Heritability quantifies the relative importance of genetic and environmental factors inexplaining the distribution of a trait or disease within a population[16]. Both kidney function and related blood biomarkers have been shown to be heritable[11,17,18]. Heritability estimates of eGFR range from 33% to 67.3%[11,18-23]. For the kidney related biomarkers, serum urea and uric acid, the heritability estimates range from 22% to 54%[18,21,24,25], and from 29% to 35%[21,25], respectively. Among subjects with type 2 diabetes, urinary albumin-creatinine ratio (UACR) shows evident familial clustering with a heritability estimate of 46%[26,27], which is higher than other population-based family studies (23% to 25%)[18,21]. The heritability of related biomarkers, such as serum electrolytes has been estimated to be moderate to high, varying from 33% to 61% for calcium[17,18,28], and 12% to 56% for

sodium and potassium[17,18,24,25,28]. To date, the heritability estimates for kidney traits have originated from twin studies[17,20,23,24,28], from family studies with relatively small sample size [11,18], or from studies in isolated founder or disease populations[21,25]. Due to sampling error, the results of these studies may have limited generalizability[16]. Therefore, heritability estimates from a large, representative sample of the general population are needed.

In this study, our aim was therefore to quantify the familial aggregation of CKD and to obtain heritability estimates of kidney traits and related biomarkers in the general population.

METHODS

Study design and population

In this cross-sectional family study, we used baseline data from the Lifelines Cohort Study and Biobank, a multidisciplinary prospective population-based cohort study of the Northern Netherlands with a unique three-generation design, and that included 167,548 subjects. It employs a broad range of investigative procedures in assessing the socio-demographic, biomedical, physical, behavioral and psychological factors which contribute to the health and disease of the general population, with a special focus on multi-morbidity and complex genetics. The overall design and rationale of this study have been described in detail elsewhere[29,30]. The recruitment of the Lifelines study was family-based by design. Eligible subjects between 20 and 50 years old were invited to participate through their general practitioner. Individuals were not invited when the participating general practitioner considered the patient not eligible, i.e. if they had severe psychiatric or physical illness, limited life expectancy or insufficient knowledge of the Dutch language. After the inclusion of these individuals their partner, children, parents and partner's parents were also invited to participate in the study. In addition, single individuals could register for participation online. In this way a three-generation family study was realized. We used the information on family members as well as information on (anonymized) names and birth dates of parents provided by all participants in questionnaires to define family relationships in Lifelines. After signing informed consent, participants received a baseline questionnaire and an invitation to a health assessment at one of the Lifelines research sites.

The Lifelines Cohort Study is conducted according to the Principles of the Declaration of Helsinki and in accordance with the research code of University Medical Center Groningen, and was approved by its medical ethical committee. All participants gave written informed consent.

Measurements

Kidney outcomes

Participants aged 8 years and older were invited to one of 12 local research sites in the north of The Netherlands for the physical examination. The baseline assessment consisted of two visits. During the first visit (duration 60 min) physical examinations were performed by a trained research nurse and containers for collection of a 24-h urine sample (age≥18 years) were handed out accompanied by oral and written instruction on how to collect this sample. Approximately 2 weeks after the first visit, a second visit (duration 10 min) was arranged to collect a fasting blood sample (age≥8 years) and hand in the collected 24-h urine.

Measurement of serum creatinine was performed by an IDMS-traceable enzymatic method on a Roche Modular analyzer using reagents and calibrators from Roche (Roche Diagnostics, Mannheim, Germany), with intra- and inter-assay coefficients of variation of 0.9% and 2.9%, respectively. Urinary albumin (UA) concentration was measured by nephelometry with a lower threshold of detection of 2.3 mg/l and intra- and inter-assay coefficient of variation of 2.2% and 2.6%, respectively (Dade Behring Diagnostic, Marburg, Germany). UA concentration was multiplied by urine volume to obtain a value of UA excretion (UAE) in milligram per 24 hours. Urinary albumin-creatinine ratio (UACR) was estimated by urinary albumin divided by urinary creatinine as measured in spot urine (age≥8 years). After addition of a constant of 1 to handle zero-values, UAE and UACR were transformed by their natural logarithm to approximate a normal distribution prior to statistical analyses.

CKD was defined as eGFR <60mL/min/1.73m^2 (CKD_{Scr}) in the complete sample. In a subsample where urinary albumin was available, we applied an additional definition of CKD according to Kidney Disease: Improving Global Outcomes (KDIGO) guidelines[31,32] (CKD_{KDIGO}, eGFR <60mL/min/1.73m^2 or UAE ≥30mg/24h, or UACR ≥30mg/g).

We calculated eGFR according to the 2012 Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) creatinine equation[33] for adults, and the Bedside Schwartz equation[34] for children (aged<18 years).

Kidney related biomarkers

The kidney related biomarkers (serum uric acid, urea and electrolytes) were measured using standard methods, i.e. for uric acid an enzymatic colorimetric assay, for urea an ultraviolet kinetic assay on a Roche Modular and for serum electrolytes (calcium, potassium, and sodium) using a Roche Modular P chemistry analyzer (Roche, Basel, Switzerland).

Covariates

Known CKD risk factors (body mass index [BMI], hypertension, type 2 diabetes, hypercholesterolemia, smoking status) were included as covariates and assessed at baseline. Blood pressure was measured ten times during 10 min with a Dinamap, PRO 100V2. The blood pressure registered was calculated by averaging the final three readings. Hypertension was defined as systolic blood pressure (SBP) ≥140 mmHg, and/or diastolic blood pressure (DBP) ≥90 mmHg, and/or self-reported prescribed use of antihypertensive drugs. Participants were categorized as having Type 2 diabetes mellitus (T2DM) if they had a measured fasting plasma glucose (FPG) ≥7.0 mmol/L, and/or a measured glycated hemoglobin (HbA1c) ≥6.5% (48 mmol/mol),[32] and/or self-reported T2DM in combination with self-reported medication use (i.e. ATC codes A10A and A10B). Hypercholesterolemia was defined as a total cholesterol of) ≥6.21 mmol/L or self-reported use of lipid-lowering drugs (ATC codes C10A, C10B). Smoking status was assessed by questionnaire and coded as smoker vs non-smoker.

Statistical analysis

Baseline characteristics

Baseline characteristics were examined for the total population. Multivariable linear regression, and multivariable logistic regression were used to examine age-adjusted differences between males and females (separately in children).

Recurrence risk ratio

The mean eGFR and prevalence of CKD were calculated for the general population and for individuals with affected first-degree relatives. Recurrence risk ratios (RRR) of CKD were calculated as the adjusted prevalence ratios between first-degree

relatives of an individual with CKD and the general population. The RRR estimated in this study is the recurrence risk ratio according to the Risch definition[35], which is the prevalence ratio between individuals with a specific type of affected relative and the general population. We used a Cox proportional hazards models, adapted according to Breslow[36], to estimate prevalence ratios in a cross-sectional study by applying an equal follow-up time for all subjects. This method has been proven to produce consistent estimates for prevalence ratios close to true limits.[37,38] A marginal proportional hazards model was used in this study to handle correlation between observations due to familial clustering. This model estimates the mean population hazard function and uses a robust sandwich method to estimate the confidence interval (CI)[39,40]. This approach has been applied and validated in previous studies on other diseases.[41-43] We calculated RRR for individuals with an affected first-degree relative of any kinship and also for individual kinship (parent, offspring, and sibling). To explore whether familial risk depends on type of kinship and sex of the affected relative, we created separate models based on type of kinship and sex of affected relatives (i.e. father, mother, son, daughter, brother, and sister). Additionally, we estimated RRR for individuals with an affected spouse (husband or wife) to quantify the effect of shared environment and/or assortative mating. In each model, we compared the risk for CKD in individuals with affected first-degree relatives or spouse with the risk in the general population. The RRR was adjusted for age, sex, BMI, hypertension, T2DM, hypercholesterolemia, and smoking. To examine the consistency of our estimates of familial recurrence risk for CKD_{SCr}, we compared results in the full sample with those in the subsample of approximately 60,000 individuals for this CKD definition.

Heritability estimates

For all continuous traits we estimated the heritability. Heritability in the narrow sense is defined as the ratio of the additive genetic variance, which reflects transmissible resemblance between relatives, to the total phenotypic variance. To estimate heritability, we used the Residual Maximum Likelihood-based variance decomposition method implemented in ASReml software[44], in which the overall phenotypic variance is decomposed into genetic and environmental components. We also included household or spousal effects in the model to estimate the influence of shared environment by using family id or spouse id as a proxy. This allowed us to distinguish between shared genes and shared environment as potential sources of familial resemblance. In addition we calculated spouse

correlations for all continuous traits. The significance of heritability was determined by using the likelihood ratio test. This test compares the likelihood of a model in which heritability is estimated, to that of a model in which heritability is constrained to zero. Age, age^2, sex, BMI, hypertension, T2DM, hypercholesterolemia, and smoking status were included as covariates in the model, irrespective of their statistical significance; we report the percentage of variance explained by these covariates (PVC). To assess the consistency of heritability estimates for eGFR, serum creatinine, and serum potassium, we compared results in the full sample with those in the subsample of approximately 60,000 participants for these traits.

All analyses were performed using ASReml (Release 4.1)[44] and R3.3.1[45]. Two-sided significance level for analyses were set at α=0.05.

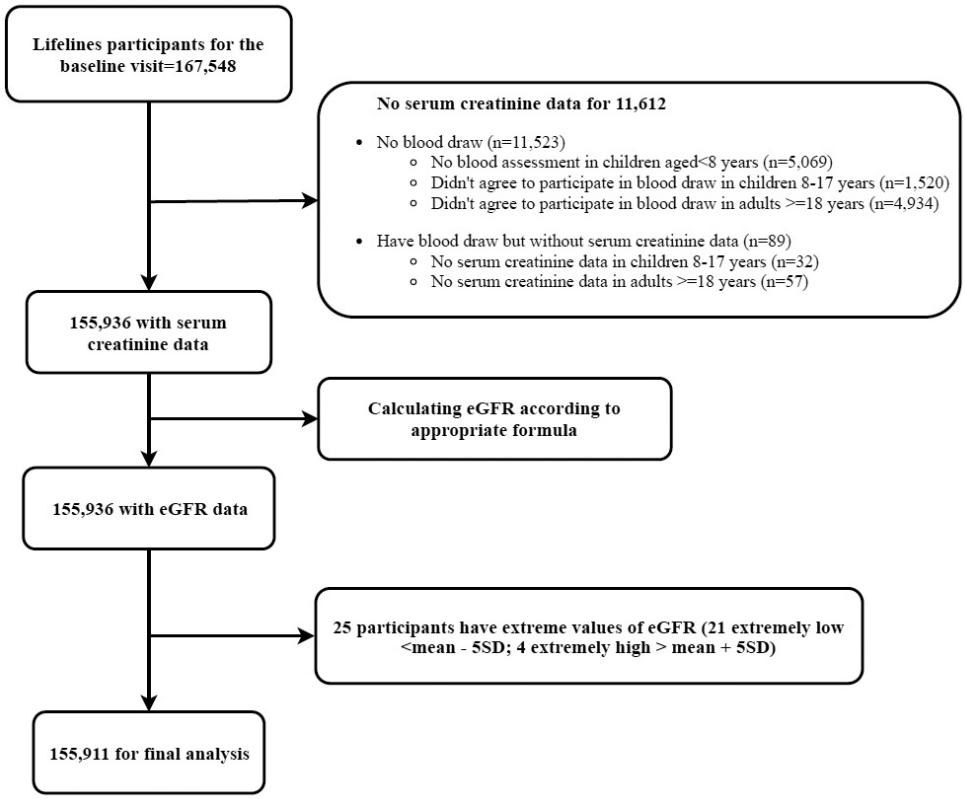
RESULTS

Baseline characteristics

We included 155,911 participants with serum creatinine and eGFR data during the baseline visit (**Figure 1**). In this full sample CKD was defined as an eGFR <60 mL/min/1.73 m^2 (CKD_{SCr}). In a subsample of approximately 60,000 participants, UAE or UACR were measured of whom 59,938 (including 743 children) had both eGFR and UACR, while 59,145 (only adults) had both eGFR and UAE data (**Supplementary Figure S1-S2**). In this subsample, CKD was defined as an eGFR <60 mL/min/1.73m^2, a UAE ≥30 mg/24h (or UACR≥ 30mg/g) or both, according to the 2011 revised Kidney Disease: Improving Global Outcomes guidelines (CKD_{KDIGO}). In the full sample there were 29,703 families (of size >=2) with an average family size of 3.92, and 39,836 singletons (i.e., individuals without any relative in the sample). The largest family consisted of 172 participants. Spouses without children were considered as a family of size 2. For the subsample, 11,477 families remained with an average family size of 3.39 and 18,537 singletons. The largest family in the subsample connected 75 participants.

Included were up to 155,911 participants (58.1% female; mean age ± SD: 43.1 ± 14.7 years) with a mean (SD) eGFR of 97.2 (15.7) mL/min/1.73 m^2 in the full sample. In the subsample of up to 59,943 subjects in which albuminuria was measured a median (interquartile range [IQR]) UAE of 3.86 (2.33-6.92) mg/24h, and a median (IQR) UACR of 2.72 (1.58-7.33) mg/g (**Table 1**) was observed. Sex stratified characteristics for adults and children revealed a slightly less favorable renal

Figure 1. Flowchart of eGFR analysis. Please note that 21 subjects with extremely low eGFR values (< 5SD from the mean) were considered CKD patients and retained in analyses of CKD. Abbreviations: estimated glomerular filtration rate, eGFR; standard deviation, SD.



risk profile for males compared to females (i.e. higher prevalence of smoking, hypertension, diabetes, and high cholesterol) but similar distributions in CKD risk and kidney markers (i.e. eGFR and UAE) (**Supplementary Table S1**). Distributions of age, sex and covariates in the subsample were similar to those in the full sample (**Supplementary Table S2**).

We identified 1862 CKD_{SCr} cases, which resulted in a crude prevalence of 1.19% (**Table 1**). A total of 2211 individuals had at least one first-degree relative with CKD_{SCr}: 1680 with at least one affected parent, 56 with at least one affected offspring, 499 with at least one affected sibling.

There was a steep increase in CKD_{SCr} prevalence after age 60. Mean eGFR was lower at higher age, and age-specific mean values of eGFR were lower among

Table 1. Baseline characteristics of adult and child participants

.	Adults (aged ≥18)	Children (aged <18)	Total
Total population, n	147715	8196	155911
Age (years)	44.83 (13.12)	12.21 (2.76)	43.11 (14.71)
Males (%)	61512 (41.64)	3891 (47.47)	65.403 (41.95)
BMI (kg/m²)	26.07 (4.33)	18.91 (3.18)	25.69 (4.56)
Current smoker (%)	31156 (21.38)	NA[a]	NA[a]
Hypertension (%)	38605 (26.13)	NA[a]	NA[a]
Diabetes (%)	5673 (3.84)	NA[a]	NA[a]
Hypercholesterolemia (%)	29888 (20.23)	NA[a]	NA[a]
Serum potassium (mEq/L)	3.86 (0.30)	3.84 (0.28)	3.86 (0.3)
Serum creatinine (mg/dL)	0.83 (0.14)	0.61 (0.13)	0.82 (0.15)
eGFR (mL/min/1.73 m²)	96.41 (15.29)	111.46 (16.81)	97.2 (15.74)
CKD_{SCr}: eGFR<60 (%)[b]	1858 (1.26)	4 (0.05)	1862 (1.19)
Subsample, n	59195	748	59943
Serum calcium (mg/dL)	9.14 (0.32)	9.50 (0.28)	9.14 (0.32)
Serum sodium (mmol/L)	141.74 (1.84)	141.63 (1.68)	141.73 (1.84)
Uric acid (mg/dL)	4.88 (1.18)	4.37 (1.01)	4.88 (1.18)
Serum urea (mg/dL)	14.48 (3.56)	12.41 (2.75)	14.45 (3.56)
UACR (mg/g)	2.72 (1.57-5.05)	2.93 (1.84-4.84)	2.72 (1.58-7.33)
UAE (mg/24h)	3.86 (2.33-6.92)	NA[a]	NA[a]
UACR ≥30	1622 (2.73)	20 (2.68)	1642 (2.73)
UAE ≥30	2431 (4.10)	NA[a]	NA[a]
CKD_{KDIGO}: eGFR<60 or UACR ≥30	3338 (5.52)	24 (3.20)	3362 (5.49)
CKD_{KDIGO}: eGFR<60 or UAE ≥30	4127 (6.83)	NA[a]	NA[a]

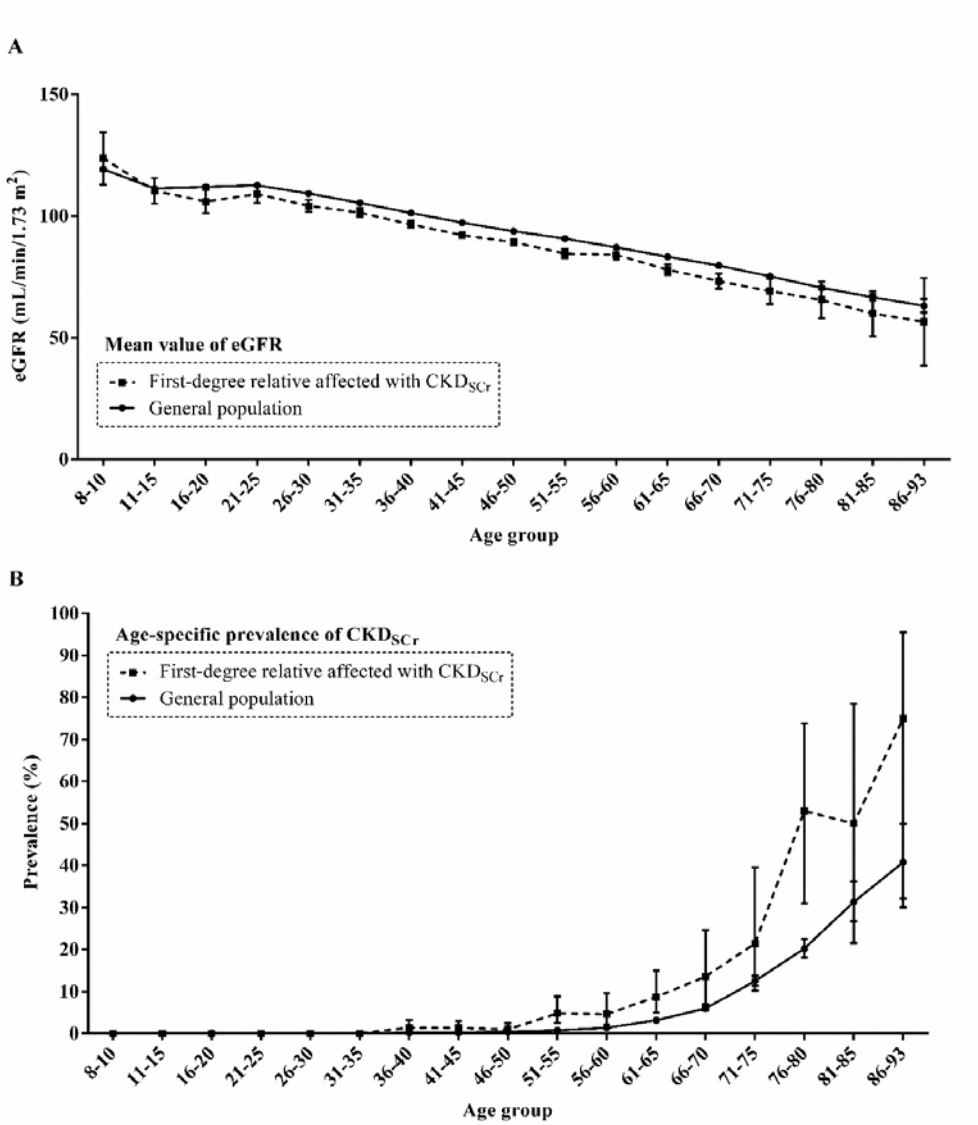
Data are presented as mean (SD), median (interquartile range) or number (%), where appropriate. Abbreviations and definitions: BMI, body mass index; SCr, serum creatinine; UACR, urinary albumin-creatinine ratio; UAE, urinary albumin excretion; eGFR, estimated glomerular filtration rate; CKD, chronic kidney disease. Conversion factors for calcium in mg/dL to mmol/L, ×0.2495; creatinine in mg/dL to µmol/L, ×88.4; uric acid in mg/dL to µmol/L, ×59.48; serum urea in mg/dL to mmol/L, ×0.357. [a]: data was not available for children. [b]: these include 21 subjects with extremely low eGFR values (< 5SD from the mean) were considered CKD patients and retained in analyses of CKD recurrence but excluded from heritability analyses

individuals with affected first-degree relatives compared to the general population (**Figure 2A**). Accordingly, the age-specific prevalence rates were also significantly higher in those with an affected first degree relative with CKD_{SCr} (**Figure 2B**). In the subsample, the crude prevalence of CKD_{KDIGO} was 5.5%-6.8% (depending on use of UACR or UAE, respectively, **Table 1**).

Recurrence risk ratio for CKD_{SCr} and CKD_{KDIGO} in individuals with affected first-degree relatives or spouses

Stratified analyses of the recurrence risk ratios (RRR) for CKD_{SCr} among individuals

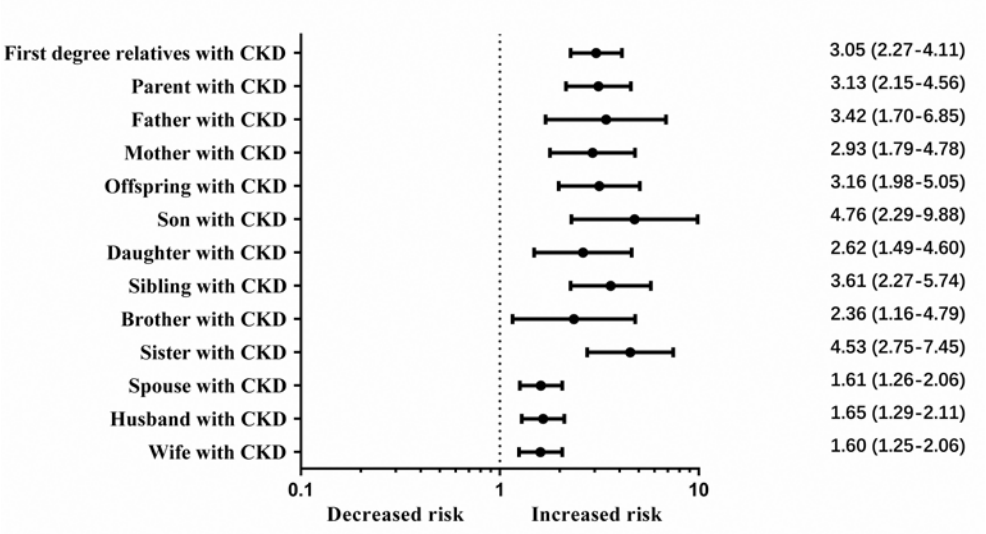
Figure 2. Comparisons of (A) age-specific mean values of eGFR and (B) age-specific prevalence of chronic kidney disease (CKD_{SCr}: eGFR<60 mL/min/1.73 m²) between individuals with affected first-degree relatives and the general population. Error bars indicate 95% confidence interval (CI).



with different affected first-degree relatives are shown in **Figure 3**. In general, having an affected first-degree relative with CKD_{SCr} was associated with an RRR of 3.05 (95% confidence interval CI: 2.27-4.11). The RRRs for CKD_{SCr} were 3.13 (95% CI: 2.15-4.56) for parents with disease, 3.16 (95% CI: 1.98-5.05) for offspring, and 3.61 (95% CI, 2.27-5.74) for siblings, respectively. Spouses of an affected individual were also at an increased risk compared to the general population (RRR = 1.61,

95% CI: 1.26-2.06). Familial recurrence showed no clear dependence on sex of the affected family member (**Figure 3**).

Figure 3. Recurrence risk ratios (adjusted for age, age2, sex, BMI, hypertension, diabetes, high cholesterol, and smoking status) for chronic kidney disease (CKD_{SCr}: eGFR<60 mL/min/1.73 m^2) in individuals with affected first degree relatives or spouse. Error bars indicate 95% confidence interval (CI).



RRRs for CKD_{SCr} in the subsample of ~60,000 participants in which albuminuria was measured showed a highly similar pattern, but values were slightly lower. The overall RRR for first degree relatives of patients with CKD_{SCr} reduced from 3.05 to 2.40.

In this subsample, prevalence of CKD_{KDIGO} was higher than that of CKD_{SCr} (**Supplementary Figure S3**). A trend was found of higher risk of CKD_{KDIGO} in first-degree relatives compared to risk in the general population, although this was less pronounced compared to the trend observed for CKD_{SCr} (**Supplementary Figure S4A and 4B**). In the subsample the RRR for first degree relatives for CKD_{KDIGO} was 1.38 (95% CI, 1.17-1.62), whereas for CKD_{SCr} this was 2.40 (95%CI, 1.78-3.23) as mentioned above (**Supplementary Figure S5**). Use of UACR instead of UAE as measure of albuminuria yielded highly similar results (**Supplementary Figures S3-S5**).

Heritability estimates

In **Table 2**, we report heritability estimates of the CKD defining traits, eGFR (44%), UAE (20%), and UACR (19%), for the kidney biomarkers serum urea (31%),

Table 2. Heritability of renal traits and related biomarkers

Traits	N	Model 1		Model 2	
		h^2 ± SE	PVC	h^2 ± SE	PVC
eGFR	155,911	0.435 ± 0.007	0.420	0.436 ± 0.007	0.423
ln(UAE) [a]	59,145	0.199 ± 0.014	0.026	0.193 ± 0.014	0.048
ln(UACR) [a]	59,938	0.185 ± 0.014	0.083	0.178 ± 0.014	0.103
Uric acid [a]	58,519	0.481 ± 0.013	0.338	0.497 ± 0.013	0.442
Serum creatinine	155,911	0.373 ± 0.007	0.374	0.379 ± 0.007	0.377
Serum urea [a]	58,481	0.307 ± 0.013	0.218	0.307 ± 0.013	0.219
Serum potassium	155,842	0.279 ± 0.007	0.041	0.278 ± 0.007	0.050
Serum calcium [a]	58,488	0.268 ± 0.013	0.059	0.266 ± 0.013	0.079
Serum sodium [a]	58,444	0.217 ± 0.013	0.066	0.221 ± 0.013	0.074

Model 1: adjusted for age, sex, age^2.
Model 2: adjusted for age, sex, age^2, body mass index, diabetes, hypertension, high cholesterol and smoking status.
Abbreviations and definitions: eGFR, estimated glomerular filtration rate; UAE, urinary albumin excretion; UACR, urinary albumin-creatinine ratio; h^2, heritability; SE, standard error; PVC, proportion of variance due to covariates.

[a] data was only available for a subsample of adult participants

serum creatinine (37%) and uric acid (48%), and finally for the serum electrolytes sodium (22%), potassium (28%), and calcium (27%). In the subsample of ~60,000 participants with available urinary albumin measurements, heritability estimates of eGFR, serum creatinine, and serum potassium were consistent with the estimates in the full sample but were less precise (**Supplementary Table S3**). Heritability estimates did not change when taking household or spousal effects into account as they explained less than 0.1% of the variance in all variables and confirmed by the small spousal correlations (**Supplementary Table S4**). Inclusion of additional covariates in model 2 did not substantially change the estimates of heritability. Around 42% of the phenotypic variance of eGFR could be explained by sex and age, whereas sex and age only explained 2.6% for UAE and 8.3% for UACR. Inclusion of additional covariates in model 2 only slightly increased the proportion of total phenotypic variance attributable to covariates (PVC) for most traits with the exception of uric acid which showed a substantial increase in PVC of 10.3%. The PVC for model 2 ranged from 4.6% for potassium to 44.1% for uric acid.

A number of traits (i.e. UAE, UACR, uric acid, serum urea, serum calcium, and serum sodium) were only available for the subsample, whereas eGFR, serum creatinine, and serum potassium were available for nearly all participants. To assess potential bias due to missingness, we repeated the heritability analysis for eGFR, serum creatinine, and serum potassium, in the subsample. Estimates were highly

comparable to those in the full sample, although the heritability estimate for serum creatinine was slightly higher in the subsample (**Supplementary Table S3**).

DISCUSSION

In this large population-based family study, we investigated the familial aggregation of CKD by comparing the risk of CKD in individuals with an affected first-degree relatives to that in the general population. Participants with an affected first-degree relative were observed to have a threefold higher risk of CKD when compared to the risk in the general population, independent of BMI, hypertension, type 2 diabetes, hypercholesterolemia, and smoking status. This may in part be explained by shared environmental factors and/or assortative mating, given that we observed a 1.6 fold higher risk in those with an affected spouse. Furthermore, we estimated the heritability of eGFR and albuminuria, as well as that of related biomarkers and electrolytes. The heritability estimate for eGFR was considerable (44%), whereas heritability of UAE was moderate (20%). Heritability of kidney related markers and serum electrolytes ranged between 20 and 50%. These results indicate an important role for genetic factors in modulating susceptibility to kidney disease in the general population.

In this study, a threefold higher risk of CKD was observed for participants with an affected first-degree relative. Previous studies that examined familial aggregation of CKD focused on its end-stage, i.e. ESKD. In African-Americans, the presence of a first-degree relative with ESKD conveyed a nine-fold increase in risk of ESKD[46], while in Taiwanese Han-Chinese, there was a 2.5-fold increase in risk[12]. In the US, a multi ancestry (African and European) population-based case-control study conducted by Lei et al. demonstrated familial aggregation of ESKD, with estimates for recurrence risk ranging from a 1.3-fold to over a tenfold increase, depending on number of family members affected[8]. In a large registry-based study among Norwegians, individuals with an affected first-degree were at a 7.2-fold higher relative risk of ESKD[15]. Among incident dialysis patients in the ESKD Network 6, 23% of subjects have close relatives with ESKD[47], and individuals with family history of ESKD are at increased risk for CKD[48]. These studies focused on ESKD as determined through registry data. Early stages of CKD remain unrecognized in such data, leading to potential misestimation of familial clustering of CKD. The present study is unique in that it included a large population and that it is based on objective laboratory

measurements of eGFR and UAE in three-generational data. Our study is therefore more sensitive to non-symptomatic, early-stage CKD in multiple family members.

The RRR of CKD_{KDIGO} in those with an affected relative was statistically significant, but considerably lower than that of CKD_{Scr}. This may be explained by our observation that measures of albuminuria were only moderately heritable, i.e. genetic factors contribute relatively little to between-individual variation in urinary albumin excretion.

Spouses of those affected by CKD were at a 1.6 times higher risk of CKD, and in addition, kidney traits showed weak but significant positive correlations between spouses. As spouses are unrelated, the increased risk of CKD in spouses and weak spousal correlations of kidney traits may reflect effects of shared environmental factors or assortative mating. To further assess the effects of shared environment on the continuous kidney traits, we examined family and spouse effects as variance components in our heritability models. As these effects were negligible, the elevated risk in spouses seems therefore more related to assortative mating, i.e. partner selection based on phenotypes that convey higher risk of CKD. In literature, strong evidence of assortment exists for factors such as substance use (e.g. smoking, alcohol use)[49], anthropometrics (e.g. height, BMI, waist-to-hip ratio), and educational attainment[50], each a potential determinant of CKD risk and progression. In the present study however, spousal correlations of eGFR and UAE did not diminish after adjustment for renal risk factors (including BMI and smoking status). Thus, assortment likely occurs on factors other than those that select for currently known CKD determinants. Future study in spousal pairs may further investigate the mechanisms and the impact of assortative mating in CKD risk.

Between-study comparison of heritability estimates is not straightforward, as phenotypic variance and contribution of genetic factors depend on population, ethnicity, environment, measurement methods, and sampling error. Some inconsistency in estimates can therefore be expected. To date, the heritability of eGFR has been described in several twin studies and a few community-based studies. In the present large-scale study, we observed a heritability of 44% for eGFR (estimated by CKD-EPI equation for serum creatinine), corroborating estimates from previous, relatively small-scale, population-based studies, such as from

Switzerland (N=1128, 46%)[18] and from South-Tyrol (N=4373, 39%)[13]. Inconsistencies in heritability estimates for eGFR can be observed with studies in specific populations or that applied different methods. A previous analysis, in pedigree data (N=1224) from the population-based Framingham Heart Study, reported a lower heritability estimate for eGFR (33%)[51]. MacCluer et al. also found a lower heritability estimate of eGFR (33%) among Zuni Indians[21]. The lower estimates in these two studies are possibly due to differences between populations, random sampling error, or use of older, less accurate eGFR estimating methods[52]. As generally observed for most traits, also for eGFR, twin studies (50%-67.3%)[20,22,23] yielded somewhat higher heritability estimates than family-based studies (33%-46%)[13,18,19,21].

The heritability estimate of urinary albumin (i.e. UAE and UACR) in the present study (20%) was similar to that in a previous Swiss population-based study by Moulin et al (23%),[18] a study that was also based on 24-hour urine collections. The heritability of UACR was 21% in Pima Indians[45], and 25% in Zuni Indians[21]. In a twin study, the heritability of UACR was 45.2±7.4%[53]. Previous studies in diabetic patients have reported highly variable heritability estimates of albuminuria measured in spot urine samples, ranging from 21% to 46%[26,27,54-56]. Finally, we observed a 22%-28% heritability for the serum electrolytes, potassium, calcium, and sodium, which confirms the potential for identifying genetic variants involved in electrolyte homeostasis in the general population.

The heritability estimates in the present study provide an upper bound to the amount of phenotypic variance that can be attributed to genetic factors. A popular method of identifying genetic factors associated with disease and disease traits is the genome-wide association study (GWAS). Large-scale GWASs have thus far identified 306 common (i.e. with minor allele frequency >1%) single nucleotide polymorphisms (SNPs) for eGFRcrea explaining 7.1% of phenotypic variance[13], whereas the present study estimates the heritability of eGFRcrea to be 44%. Similarly, the 59 SNPs thus far identified in GWASs on UACR explained 0.7%, which is modest compared to our heritability estimate of 20%[14]. Thus, much of the heritability of kidney traits remains to be discovered. Potentially, future whole genome sequencing study that focus on rare variants (i.e. SNPs with a minor allele frequency <1%) may unveil a large proportion of this missing heritability[57,58].
The present study is by far the largest family-based study on kidney traits that

uses laboratory defined CKD, whereas other similar studies relied on medical records or health insurance data on ESKD that are not sensitive to earlier stage CKD. Leveraging these laboratory data, the present study is the first to quantify the familial clustering of CKD including early (i.e. non-ESKD) stages of CKD. Furthermore, Lifelines is representative with regards to its source population (i.e. the general population of the northern Netherlands), which facilitated precise estimation of heritability. In addition, albuminuria was determined not only in spot urine samples, but also in 24h urine collections, that are considered the gold standard to assess albuminuria. There are only very few large scale epidemiological studies that have 24h urine collections available. .

Several limitations need to be addressed. First, although gold standard 24h albuminuria measurements were available, this was only true for a subsample of approximately 60,000 participants. However, this substantial sample size still offers ample statistical power and reliable results. Furthermore, missingness was likely random as age, sex and covariate distributions were highly similar between sub- and full sample, and therefore unlikely to have seriously biased our results. This is supported by the consistency of the heritability estimates of eGFR, serum creatinine, and serum potassium between the total sample and the subsample. Second, GFR was not measured directly but estimated from serum creatinine. Therefore, some bias is possible due to creatinine metabolism. In addition, eGFR estimating equations are known to be less precise in the higher ranges (>60 mL/min/1.73m^2)[59-61]. These measurement errors may have caused downward bias in our heritability estimates. Third, no kidney biopsy data was available, nor could we exclude Mendelian forms of inherited kidney disease; we therefore could not distinguish between the different etiologies of CKD. Fourth, potential preferential missingness of data from non-participating affected family members may have led to underestimation of recurrence risk ratios. Finally, 98% of Lifelines participants are of European ancestry[62]; we therefore cannot generalize our results to other ancestries.

The results of this study may have several implications in addition to those previously mentioned. First, the data on familial recurrence of CKD may guide clinical decision-making with regards to CKD diagnosis and prevention. Further study is warranted to assess the added value of family history in risk stratification of CKD, and to investigate the potential impact of specifically targeting family members of CKD patients for screening and prevention strategies. Second, the

heritability estimates provide an upper bound to how much variance of a trait can be explained by genetic factors. Future studies, e.g. GWAS, may focus on identifying these genetic factors.

In summary, we demonstrate that CKD clusters in families in the general population, given that risk of CKD was strongly elevated in those with an affected relative. Considerable heritability (20-50%) of kidney traits was observed. Therefore, much of the familial clustering may be attributed to genetic factors. The data presented in this study inform future work on risk stratification based on family history, and provide a step forward in disentangling genetic and environmental risk factors in CKD.

ACKNOWLEDGEMENTS

We acknowledge the services of the Lifelines Cohort Study, the contributing research centers delivering data to Lifelines, and all the study participants. We also appreciate the help from Dr. Chang Fu Kuo in modeling recurrence risk ratio of CKD in the general population.

Author Contributions: Study conception and design: HS, CHLT, JZ; data analysis: CHLT, JZ; data interpretation and literature review: JZ, CHLT, HS, RTG; research mentorship and supervision: HS, RTG. Each author contributed important intellectual content during manuscript drafting or revision and accepts accountability for the author's own contribution, and agrees to ensure that questions pertaining to the accuracy or integrity of any portion of the work are appropriately investigated and resolved.

Support: The Lifelines cohort study was supported by the Netherlands Organization of Scientific Research NWO (grant 175.010.2007.006), the Economic Structure Enhancing Fund (FES) of the Dutch Government, the Ministry of Economic Affairs, the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the Northern Netherlands Collaboration of Provinces (SNN), the Province of Groningen, the University Medical Center Groningen, the University of Groningen, the Dutch Kidney Foundation and the Dutch Diabetes Research Foundation. Author JZ was awarded a personal grant by China Scholarship Council (CSC) during his study period in the Netherlands.

Financial disclosure: The authors declare that they have no other relevant financial interests.

REFERENCES

1. Jha V, Garcia-Garcia G, Iseki K, et al. Chronic kidney disease: Global dimension and perspectives. *The Lancet*. 2013;382(9888):260-272.
2. Bruck K, Stel VS, Gambaro G, et al. CKD prevalence varies across the European general population. *J Am Soc Nephrol*. 2016;27(7):2135-2147.
3. Go AS, Chertow GM, Fan D, McCulloch CE, Hsu CY. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *N Engl J Med*. 2004;351(13):1296-1305.
4. Kahn MR, Robbins MJ, Kim MC, Fuster V. Management of cardiovascular disease in patients with kidney disease. *Nat Rev Cardiol*. 2013;10(5):261.
5. Gansevoort RT, Correa-Rotter R, Hemmelgarn BR, et al. Chronic kidney disease and cardiovascular risk: Epidemiology, mechanisms, and prevention. *The Lancet*. 2013;382(9889):339-352.
6. Nitsch D, Grams M, Sang Y, et al. Associations of estimated glomerular filtration rate and albuminuria with mortality and renal failure by sex: A meta-analysis. *BMJ*. 2013;346:f324.
7. Echouffo-Tcheugui JB, Kengne AP. Risk models to predict chronic kidney disease and its progression: A systematic review. *PLoS Med*. 2012;9(11):e1001344. Epub 2012 Nov 20.
8. Lei HH, Perneger TV, Klag MJ, Whelton PK, Coresh J. Familial aggregation of renal disease in a population-based case-control study. *J Am Soc Nephrol*. 1998;9(7):1270-1276.
9. Queiroz Madeira EP, da Rosa Santos O, Ferreira Santos SF, Alonso da Silva L, MacIntyre Innocenzi A, Santoro-Lopes G. Familial aggregation of end-stage kidney disease in Brazil. *Nephron*. 2002;91(4):666-670.
10. Satko SG, Freedman BI. The familial clustering of renal disease and related phenotypes. *Med Clin North Am*. 2005;89(3):447-456.
11. Fava C, Montagnana M, Burri P, et al. Determinants of kidney function in swedish families: Role of heritable factors. *J Hypertens*. 2008;26(9):1773-1779.
12. Wu HH, Kuo CF, Li IJ, et al. Family aggregation and heritability of ESRD in Taiwan: A population-based study. *Am J Kidney Dis*. 2017;70(5):619-626.
13. Wuttke M, Li Y, Li M, et al. A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat Genet*. 2019;51(6):957-972.
14. Teumer A, Li Y, Ghasemi S, et al. Genome-wide association meta-analyses and fine-mapping elucidate pathways influencing albuminuria. *Nat Commun*. 2019;10(1):1-19.
15. Skrunes R, Svarstad E, Reisaeter AV, Vikse BE. Familial clustering of ESRD in the Norwegian population. *Clin J Am Soc Nephrol*. 2014;9(10):1692-1700.
16. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era - concepts and misconceptions. *Nat Rev Genet*. 2008;9(4):255-266.
17. Hunter DJ, Lange M, Snieder H, et al. Genetic contribution to renal function and electrolyte balance: A twin study. *Clin Sci (Lond)*. 2002;103(3):259-265.
18. Moulin F, Ponte B, Pruijm M, et al. A population-based approach to assess the heritability and distribution of renal handling of electrolytes. *Kidney Int*. 2017;92(6):1536-1543.
19. Bochud M, Elston RC, Maillard M, et al. Heritability of renal function in hypertensive families of African descent in the Seychelles (Indian ocean). *Kidney Int*. 2005;67(1):61-69.
20. Raggi P, Su S, Karohl C, Veledar E, Rojas-Campos E, Vaccarino V. Heritability of renal function and inflammatory markers in adult male twins. *Am J Nephrol*. 2010;32(4):317-323.
21. MacCluer JW, Scavini M, Shah VO, et al. Heritability of measures of kidney disease among Zuni Indians: The Zuni kidney project. *Am J K Dis*. 2010;56(2):289-302.
22. Pasha DN, Davis JT, Rao F, et al. Heritable influence of DBH on adrenergic and renal function: Twin and disease studies. *PloS one*. 2013;8(12):e82956.

23. Arpegard J, Viktorin A, Chang Z, de Faire U, Magnusson PK, Svensson P. Comparison of heritability of cystatin C- and creatinine-based estimates of kidney function and their relation to heritability of cardiovascular disease. *J Am Heart Assoc*. 2015;4(1):e001467.

24. Bathum L, Fagnani C, Christiansen L, Christensen K. Heritability of biochemical kidney markers and relation to survival in the elderly—results from a Danish population-based twin study. *Clin Chim Acta*. 2004;349(1-2):143-150.

25. Pilia G, Chen W, Scuteri A, et al. Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet*. 2006;2(8):e132.

26. Forsblom CM, Kanninen T, Lehtovirta M, Saloranta C, Groop LC. Heritability of albumin excretion rate in families of patients with type II diabetes. *Diabetologia*. 1999;42(11):1359-1366.

27. Langefeld CD, Beck SR, Bowden DW, Rich SS, Wagenknecht LE, Freedman BI. Heritability of GFR and albuminuria in caucasians with type 2 diabetes mellitus. *Am J Kidney Dis*. 2004;43(5):796-800.

28. Whitfield J, Martin N. The effects of inheritance on constituents of plasma: A twin study on some biochemical variables. *Ann Clin Biochem*. 1984;21(3):176-183.

29. Stolk RP, Rosmalen JG, Postma DS, et al. Universal risk factors for multifactorial diseases. *Eur J Epidemiol*. 2008;23(1):67-74.

30. Scholtens S, Smidt N, Swertz MA, et al. Cohort profile: LifeLines, a three-generation cohort study and biobank. *Int J Epidemiol*. 2015 Aug;44(4):1172-80.

31. Levey AS, de Jong PE, Coresh J, et al. The definition, classification, and prognosis of chronic kidney disease: A KDIGO controversies conference report. *Kidney Int*. 2011;80(1):17-28.

32. Levin A, Stevens PE, Bilous RW, et al. Kidney disease: Improving global outcomes (KDIGO) CKD work group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney Int Supp*. 2013;3(1):1-150.

33. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med*. 2009;150(9):604-612.

34. Schwartz GJ, Munoz A, Schneider MF, et al. New equations to estimate GFR in children with CKD. *J Am Soc Nephrol*. 2009;20(3):629-637.

35. Risch N. Linkage strategies for genetically complex traits. I. multilocus models. *Am J Hum Genet*. 1990;46(2):222-228.

36. Breslow N. Covariance analysis of censored survival data. *Biometrics*. 1974:89-99.

37. Skov T, Deddens J, Petersen MR, Endahl L. Prevalence proportion ratios: Estimation and hypothesis testing. *Int J Epidemiol*. 1998;27(1):91-95.

38. Barros AJ, Hirakata VN. Alternatives for logistic regression in cross-sectional studies: An empirical comparison of models that directly estimate the prevalence ratio. *BMC medical research methodology*. 2003;3(1):21.

39. Lin D. Cox regression analysis of multivariate failure time data: The marginal approach. *Stat Med*. 1994;13(21):2233-2247.

40. Biswas A, Datta S, Fine JP, Segal MR. *Statistical advances in the biomedical science*. Wiley Online Library; 2007.

41. Lichtenstein P, Yip BH, Björk C, et al. Common genetic determinants of schizophrenia and bipolar disorder in swedish families: A population-based study. *The Lancet*. 2009;373(9659):234-239.

42. Kuo C, Grainge MJ, Valdes AM, et al. Familial risk of Sjögren's syndrome and co-aggregation of autoimmune diseases in affected families: A nationwide population study. *Arthritis & Rheumatology*. 2015;67(7):1904-1912.

43. Kuo C, Grainge MJ, Valdes AM, et al. Familial aggregation of systemic lupus erythematosus and coaggregation of autoimmune diseases in affected families. *JAMA Int Med*. 2015;175(9):1518-1526.

44. Gilmour A, Gogel B, Cullis B, Welham S, Thompson R. ASReml user guide release 4.1 structural specification. *Hemel hempstead: VSN international ltd*. 2015.

45. R Core Team. R: A language and environment for statistical computing. R foundation for statistical computing, vienna, Austria.

46. Freedman BI, Spray BJ, Tuttle AB, Buckalew Jr VM. The familial risk of end-stage renal disease in african americans. *Am J Kidney Dis*. 1993;21(4):387-393.

47. Freedman BI, Volkova NV, Satko SG, et al. Population-based screening for family history of end-stage renal disease among incident dialysis patients. *Am J Nephrol*. 2005;25(6):529-535.

48. McClellan WM, Satko SG, Gladstone E, Krisher JO, Narva AS, Freedman BI. Individuals with a family history of ESRD are a high-risk population for CKD: Implications for targeted surveillance and intervention activities. *Am J Kidney Dis*. 2009;53(3):S100-S106.

49. Agrawal A, Heath AC, Grant JD, et al. Assortative mating for cigarette smoking and for alcohol consumption in female Australian twins and their spouses. *Behav Genet*. 2006;36(4):553-566.

50. Robinson MR, Kleinman A, Graff M, et al. Genetic evidence of assortative mating in humans. *Nat Hum Behaviour*. 2017;1(1):0016.

51. Fox CS, Yang Q, Cupples LA, et al. Genomewide linkage analysis to serum creatinine, GFR, and creatinine clearance in a community-based population: The Framingham heart study. *J Am Soc Nephrol*. 2004;15(9):2457-2461.

52. Michels WM, Grootendorst DC, Verduijn M, Elliott EG, Dekker FW, Krediet RT. Performance of the cockcroft-gault, MDRD, and new CKD-EPI formulas in relation to GFR, age, and body size. *Clin J Am Soc Nephrol*. 2010;5(6):1003-1009.

53. Rao F, Wessel J, Wen G, et al. Renal albumin excretion: Twin studies identify influences of heredity, environment, and adrenergic pathway polymorphism. *Hypertension*. 2007;49(5):1015-1031.

54. Fogarty DG, Rich SS, Hanna L, Warram JH, Krolewski AS. Urinary albumin excretion in families with type 2 diabetes is heritable and genetically correlated to blood pressure. *Kidney Int*. 2000;57(1):250-257.

55. Imperatore G, Knowler WC, Pettitt DJ, Kobes S, Bennett PH, Hanson RL. Segregation analysis of diabetic nephropathy in Pima Indians. *Diabetes*. 2000;49(6):1049-1056.

56. Krolewski AS, Poznik G, Placha G, et al. A genome-wide linkage scan for genes controlling variation in urinary albumin excretion in type II diabetes. *Kidney Int*. 2006;69(1):129-136.

57. Nolte IM, Tropf FC, Snieder H. Missing heritability of complex traits and diseases. *eLS*. 2001:1-9.

58. Wainschtein P, Jain DP, Yengo L, et al. Recovery of trait heritability from whole genome sequence data. *bioRxiv*. 2019:588020.

59. Stevens LA, Coresh J, Greene T, Levey AS. Assessing kidney function—measured and estimated glomerular filtration rate. *N Engl J Med*. 2006;354(23):2473-2483.

60. Earley A, Miskulin D, Lamb EJ, Levey AS, Uhlig K. Estimating equations for glomerular filtration rate in the era of creatinine standardization: A systematic review. *Ann Intern Med*. 2012;156(11):785-795.

61. Levey AS, Inker LA, Coresh J. GFR estimation: From physiology to public health. *Am J Kidney Dis*. 2014;63(5):820-834.

62. van der Ende, M Yldau, Hartman MH, Hagemeijer Y, et al. The LifeLines cohort study: Prevalence and treatment of cardiovascular disease and risk factors. *Int J Cardiol*. 2017;228:495-500.

Supplementary Material

CHAPTER 4

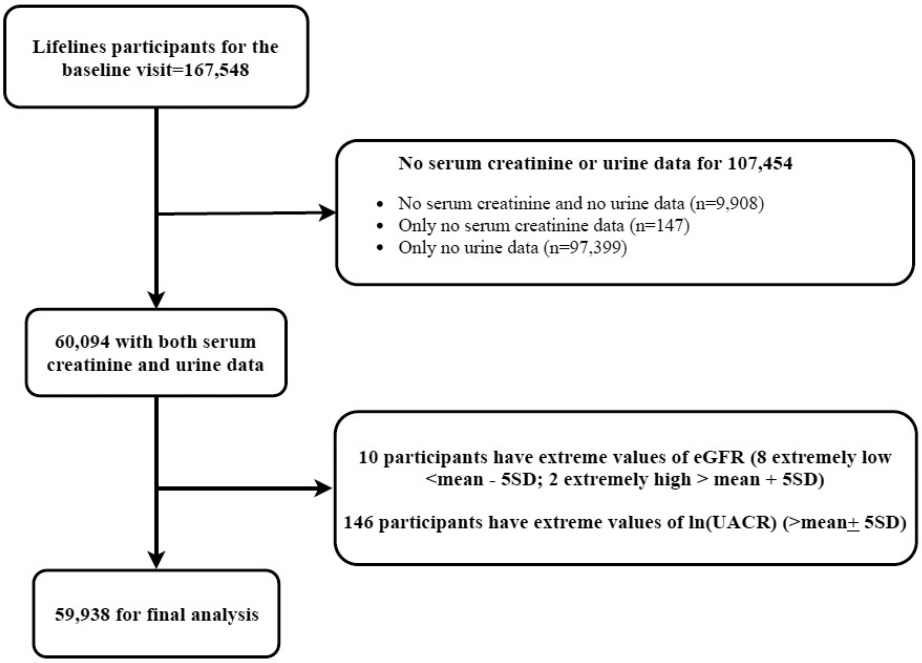


Figure S1. Flowchart of eGFR and UACR analysis. Abbreviations: estimated glomerular filtration rate, eGFR; standard deviation, SD; urinary albumin-creatinine ratio, UACR.

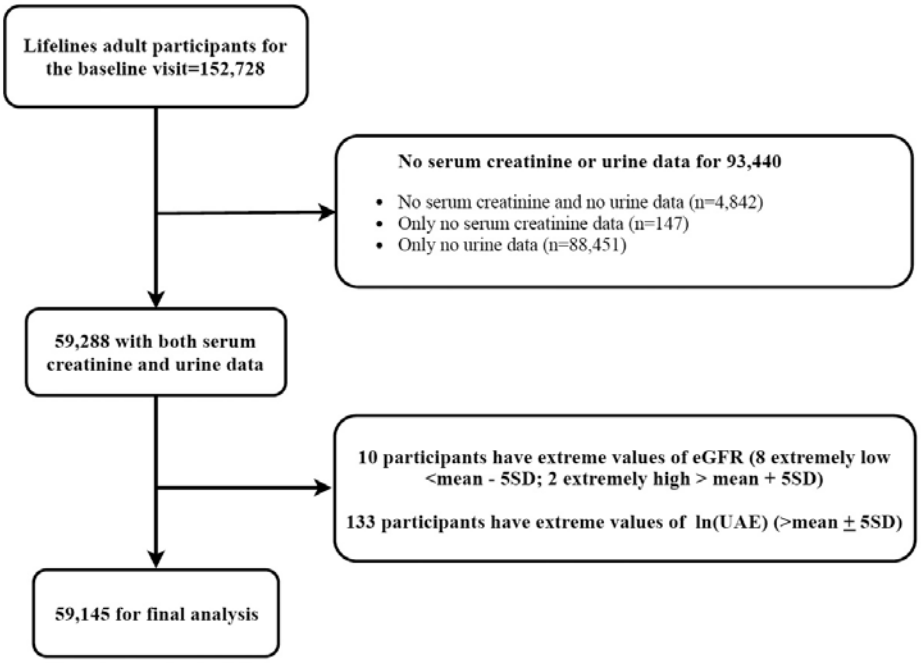


Figure S2. Flowchart of eGFR and UAE analysis. Abbreviations: estimated glomerular filtration rate, eGFR; standard deviation, SD; urinary albumin excretion, UAE.

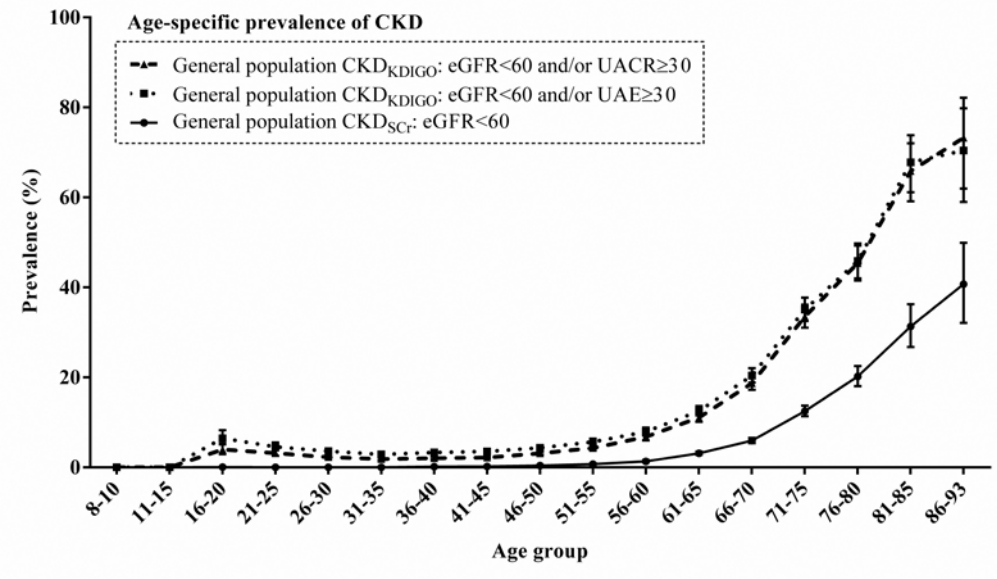


Figure S3. Age-specific prevalence of chronic kidney disease (CKD_{SCr}: eGFR<60 mL/min/1.73 m²; CKDKDIGO: eGFR < 60 mL/min/1.73 m² and/or UAE (UACR) ≥ 30 mg/24 hours (mg/g)) in the general population diagnosed by different criteria. Error bars indicate 95% confidence interval (CI). Abbreviations: estimated glomerular filtration rate, eGFR; urinary albumin excretion, UAE; urinary albumin-creatinine ratio, UACR.

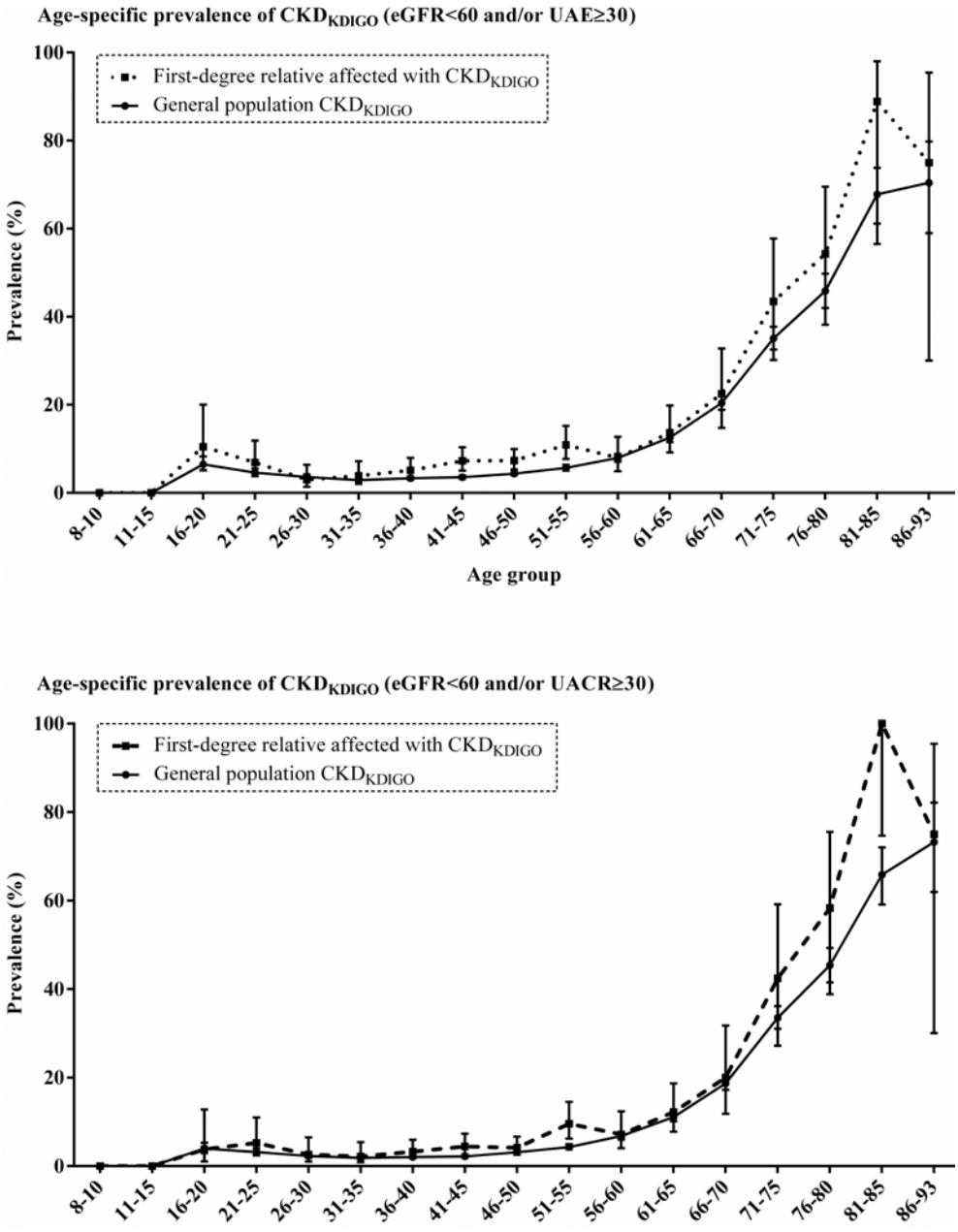


Figure S4. Comparisons of age-specific prevalence of chronic kidney disease (CKD_{KDIGO}: eGFR < 60 mL/min/1.73 m² and/or UAE (UACR) ≥ 30 mg/24 hours (mg/g)) between individuals with affected first-degree relatives and the general population. Error bars indicate 95% confidence interval (CI). Abbreviations: estimated glomerular filtration rate, eGFR; urinary albumin excretion, UAE; urinary albumin-creatinine ratio, UACR.

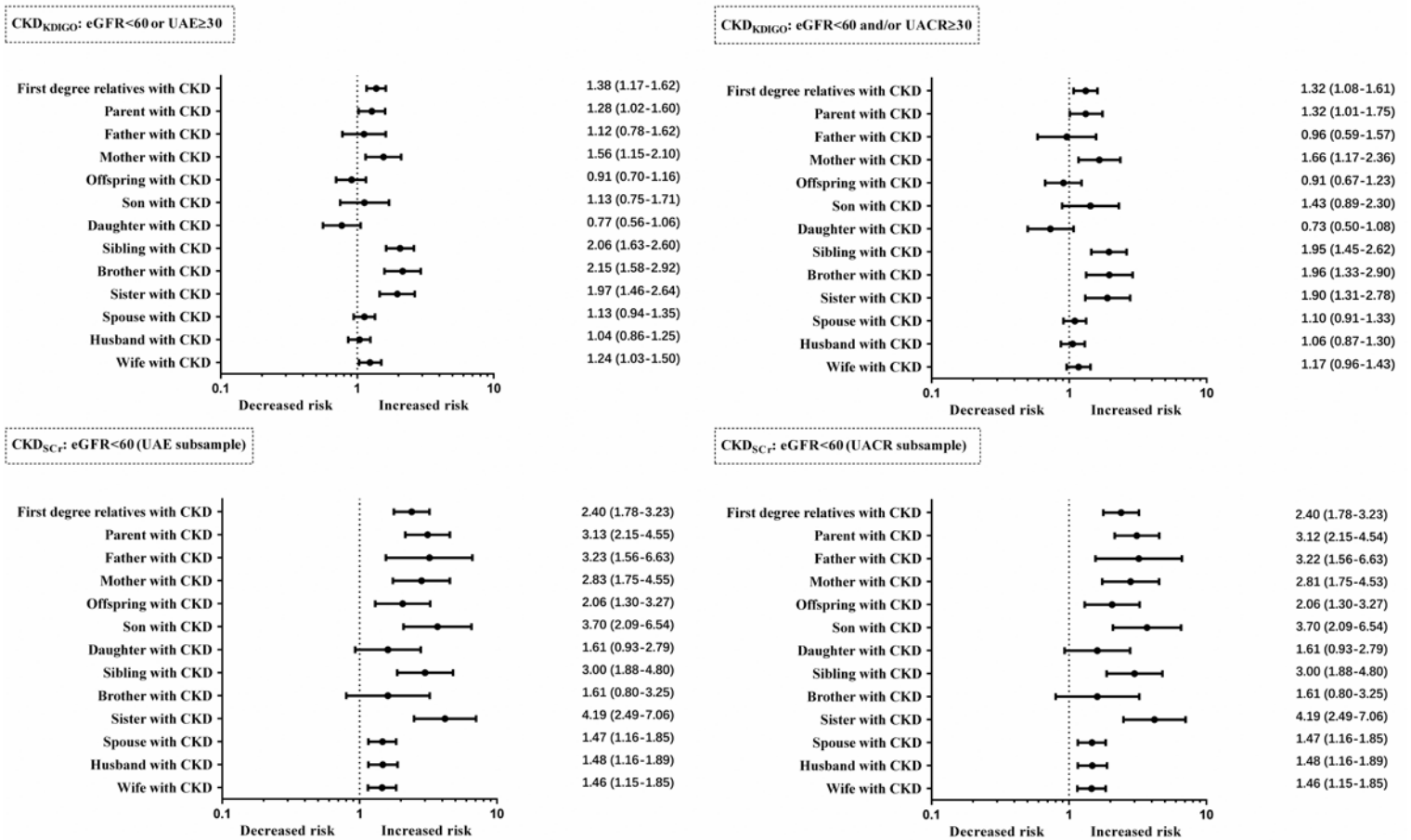


Figure S5. Recurrence risk ratios (adjusted for age, age², sex, BMI, hypertension, diabetes, high cholesterol, and smoking status) for chronic kidney disease (CKD) in individuals with affected first degree relatives or spouse. Error bars indicate 95% confidence interval (CI). Abbreviations: estimated glomerular filtration rate, eGFR; urinary albumin excretion, UAE; urinary albumin-creatinine ratio, UACR.

Table S1. Baseline characteristics of adult and child participants stratified by sex and age

Characteristics	Adults (aged ≥18)				Children (aged <18)				Total
	Total	Males	Females	P[a]	Total	Males	Females	P[a]	
Total population, n (%)	147715	61512 (41.64)	86203 (58.36)		8196	3891 (47.47)	4305 (52.53)		155911
Age (years)	44.83 (13.12)	45.46 (13.18)	44.38 (13.05)	<0.001	12.21 (2.76)	12.05 (2.72)	12.35 (2.79)	<0.001	43.11 (14.71)
BMI (kg/m²)	26.07 (4.33)	26.37 (3.68)	25.85 (4.72)	<0.001	18.91 (3.18)	18.61 (2.99)	19.19 (3.31)	<0.001	25.69 (4.56)
Current smoker (%)	31156 (21.38)	14245 (23.39)	17085 (19.94)	<0.001	NA[c]	NA[c]	NA[c]	NA[c]	NA[c]
Hypertension (%)	38605 (26.13)	19876 (32.31)	18729 (21.73)	<0.001	NA[c]	NA[c]	NA[c]	NA[c]	NA[c]
Diabetes (%)	5673 (3.84)	2805 (4.56)	2868 (3.33)	<0.001	NA[c]	NA[c]	NA[c]	NA[c]	NA[c]
Hypercholesterolemia (%)	29888 (20.23)	14361 (23.35)	15527 (18.01)	<0.001	NA[c]	NA[c]	NA[c]	NA[c]	NA[c]
Serum potassium (mEq/L)	3.86 (0.30)	3.89 (0.33)	3.84 (0.29)	<0.001	3.84 (0.28)	3.85 (0.28)	3.83 (0.28)	<0.001	3.86 (0.3)
Serum creatinine (mg/dL)	0.83 (0.14)	0.93 (0.13)	0.76 (0.11)	<0.001	0.61 (0.13)	0.61 (0.14)	0.60 (0.12)	<0.001	0.82 (0.15)
eGFR (mL/min/1.73 m²)	96.41 (15.29)	97.34 (15.22)	95.74 (15.31)	<0.001	111.46 (16.81)	111.35 (16.84)	111.56 (16.78)	<0.001	97.2 (15.74)
CKD_{SCr}; eGFR<60 (%)[d]	1858 (1.26)	806 (1.31)	1052 (1.22)	0.0352	4 (0.05)	1	3	NA[b]	1862 (1.19)
Subsample, n (%)	59195	24517 (41.42)	34678 (58.58)		748	356 (47.59)	392 (52.41)		59943
Age (years)	44.94 (12.52)	45.46 (12.60)	44.57 (12.45)	<0.001	12.27 (2.78)	12.20 (2.73)	12.34 (2.82)	0.4748	44.53 (12.96)
Serum calcium (mg/dL)	9.14 (0.32)	9.22 (0.32)	9.10 (0.32)	<0.001	9.50 (0.28)	9.54 (0.28)	9.46 (0.28)	0.0019	9.14 (0.32)
Serum sodium (mEq/L)	141.74 (1.84)	142.23 (1.76)	141.39 (1.85)	<0.001	141.63 (1.68)	141.69 (1.82)	141.57 (1.54)	0.287	141.73 (1.84)
Uric acid (mg/dL)	4.88 (1.18)	5.72 (1.01)	4.37 (1.01)	<0.001	4.37 (1.01)	4.54 (1.01)	4.20 (0.84)	<0.001	4.88 (1.18)
Serum urea (mg/dL)	14.48 (3.56)	15.77 (3.61)	13.61 (3.33)	<0.001	12.41 (2.75)	12.97 (2.69)	11.88 (2.72)	<0.001	14.45 (3.56)
UACR (mg/g)	2.72 (1.57-5.05)	2.00 (1.20-3.70)	3.29 (2.00-5.85)	<0.001	2.93 (1.84-4.84)	2.52 (1.64-4.52)	3.12 (1.97-5.41)	<0.001	2.72 (1.58-7.33)
UAE (mg/24h)	3.86 (2.33-6.92)	3.68 (2.18-6.78)	4.00 (2.44-7.01)	<0.001	NA[c]	NA[c]	NA[c]	NA[c]	NA[c]
UACR ≥30	1622 (2.73)	688 (2.80)	934 (2.69)	0.883	20 (2.68)	5 (1.41)	15 (3.85)	0.053	1642 (2.73)
UAE ≥30	2431 (4.10)	1254 (5.10)	1177 (3.39)	<0.001	NA[c]	NA[c]	NA[c]	NA[c]	NA[c]
CKD_{KDIGO}; eGFR<60 or UACR ≥30	3338 (5.52)	1400 (5.58)	1938 (5.48)	0.0325	24 (3.20)	6 (1.69)	18 (4.58)	0.0354	3362 (5.49)
CKD_{KDIGO}; eGFR<60 or UAE ≥30	4127 (6.83)	1949 (7.78)	2178 (6.16)	<0.001	NA[c]	NA[c]	NA[c]	NA[c]	NA[c]

Data are presented as mean (SD), median (interquartile range) or number (%), where appropriate. Abbreviations and definitions: BMI, body mass index; SCr, serum creatinine; UACR, urinary albumin-creatinine ratio; UAE, urinary albumin excretion; eGFR, estimated glomerular filtration rate; CKD, chronic kidney disease. Conversion factors for calcium in mg/dL to mmol/L, ×0.2495; creatinine in mg/dL to µmol/L, ×88.4; uric acid in mg/dL to µmol/L, ×59.48; serum urea in mg/dL to mmol/L, ×0.357. [a]: P-values were adjusted for age using multiple linear regression, and multiple logistic regression where appropriate. [b]: sample size was too small to calculate P-value. [c]: data was not available for children. [d]: 21 subjects with extremely low eGFR values (< 5SD from the mean) were considered CKD patients and retained in analyses of CKD

Table S2. Baseline characteristics of adult and child participants among subsample stratified by sex and age

Characteristics	Adults (aged ≥18)				Children (aged <18)				Total
	Total	Males	Females	P[a]	Total	Males	Females	P[a]	
Subsample, n (%)	59195	24517 (41.42)	34678 (58.58)		748	356 (47.59)	392 (52.41)		59943
Age (years)	44.94 (12.52)	45.46 (12.60)	44.57 (12.45)	<0.001	12.27 (2.78)	12.20 (2.73)	12.34 (2.82)	0.474	44.53 (12.96)
BMI (kg/m²)	26.08 (4.31)	26.42 (3.65)	25.84 (4.70)	<0.001	19.11 (3.17)	18.71 (2.71)	19.48 (3.50)	<0.001	25.99 (4.37)
Current smoker (%)	13090 (22.55%)	5893 (24.53)	7197 (21.16)	<0.001	NA[c]	NA[c]	NA[c]	NA[c]	NA[c]
Hypertension (%)	15951 (26.95)	8302 (33.86)	7649 (22.06)	<0.001	NA[c]	NA[c]	NA[c]	NA[c]	NA[c]
Diabetes (%)	2324 (3.93)	1162 (4.74)	1162 (3.35)	<0.001	NA[c]	NA[c]	NA[c]	NA[c]	NA[c]
Hypercholesterolemia (%)	11397 (19.25)	5444 (22.21)	5953 (17.17)	<0.001	NA[c]	NA[c]	NA[c]	NA[c]	NA[c]
Serum potassium (mEq/L)	3.90 (0.30)	3.92 (0.31)	3.88 (0.29)	<0.001	3.95 (0.29)	3.97 (0.29)	3.93 (0.28)	0.087	3.90 (0.30)
Serum creatinine (mg/dL)	0.84 (0.14)	0.94 (0.12)	0.77 (0.11)	<0.001	0.62 (0.13)	0.63 (0.14)	0.61 (0.11)	0.002	0.84 (0.14)
eGFR (mL/min/1.73 m²)	95.33 (14.92)	96.55 (14.77)	94.47 (14.97)	<0.001	108.55 (15.23)	108.31 (15.52)	108.77 (14.98)	0.377	95.49 (15.00)
CKD_{Scr}: eGFR<60 (%)[d]	709 (1.20)	306 (1.25)	403 (1.16)	0.482	0	0	0	NA[b]	709 (1.18)

Data are presented as mean (SD), median (interquartile range) or number (%), where appropriate. Abbreviations and definitions: BMI, body mass index; SCr, serum creatinine; eGFR, estimated glomerular filtration rate; CKD, chronic kidney disease. Conversion factors for creatinine in mg/dL to μmol/L, ×88.4. [a]: P-values were adjusted for age using multiple linear regression, and multiple logistic regression where appropriate. [b]: sample size was too small to calculate P-value. [c]: data was not available for children. [d]: 21 subjects with extremely low eGFR values (< 5SD from the mean) were considered CKD patients and retained in analyses of CKD

Table S3. Heritability of eGFR, serum creatinine and potassium in subsample participants only

Traits	N	Model 1		Model 2	
		h^2 ± SE	PVC	h^2 ± SE	PVC
eGFR	59.943	0.456 ± 0.013	0.387	0.458 ± 0.013	0.388
Serum creatinine	59.943	0.433 ± 0.012	0.361	0.436 ± 0.012	0.363
Serum potassium	59.943	0.262 ± 0.013	0.034	0.263 ± 0.013	0.041

Model 1: adjusted for age, sex, age^2.
Model 2: adjusted for age, sex, age^2, body mass index, diabetes, hypertension, high cholesterol, and smoking status.

Abbreviations and definitions: eGFR, estimated glomerular filtration rate; UAE, urinary albumin excretion; UACR, urinary albumin-creatinine ratio; h^2, heritability; SE, standard error; PVC, proportion of variance due to covariates.

Table S4. Spousal correlations

Renal traits	N (Pairs)	model1		model2	
		Pearson correlation	P value	Pearson correlation	P value
eGFR	29356	0.069	<0.001	0.067	<0.001
serum creatinine	29356	0.074	<0.001	0.073	<0.001
ln(UAE)	9935	0.035	<0.001	0.034	<0.001
ln(UACR)	9951	0.032	<0.001	0.034	<0.001
serum urea	9933	0.116	<0.001	0.115	<0.001
uric acid	9948	0.082	<0.001	0.050	<0.001
serum potassium	29325	0.167	<0.001	0.166	<0.001
serum calcium	9934	0.034	<0.001	0.035	<0.001
serum sodium	9913	0.077	<0.001	0.069	<0.001
Chronic Kidney Disease (CKD)	N (Pairs)	Phi coefficient		P value	
CKD_{SCr}: $eGFR_{SCr}$<60 (%)	29356	0.12		<0.001	
CKD_{KDIGO}: $eGFR_{SCr}$<60 or UACR ≥30	9951	0.15		<0.001	
CKD_{KDIGO}: $eGFR_{SCr}$<60 or UAE ≥30	9935	0.12		<0.001	

Model 1: adjusted for age, age^2, and sex
Model 2: adjusted for age, age^2, sex, BMI, hypertension, diabetes, high cholesterol, and smoking status

Abbreviations and definitions: eGFR, estimated glomerular filtration rate; UAE, urinary albumin excretion; UACR, urinary albumin-creatinine ratio

4

Evaluation of a genetic risk score based on creatinine-estimated glomerular filtration rate and its association with kidney outcomes

Chris HL Thio, Peter J van der Most, Ilja M Nolte, Pim van der Harst, Ute Bültmann, Ron T Gansevoort, Harold Snieder

Nephrol Dial Transplant. 2017;33(10):1757-64.

CHAPTER

5

ABSTRACT

Introduction. Cross-sectional GWAS on creatinine-estimated GFR ($eGFR_{crea}$) identified 53 SNPs. These SNP effects can be aggregated into a Genetic Risk Score (GRS) for chronic kidney disease (CKD). To assess its clinical utility, we examined associations with creatinine-estimated kidney outcomes, both cross-sectionally and longitudinally. Additionally, we examined associations with cystatin C-estimated kidney outcomes to verify that a GRS based on $eGFR_{crea}$ SNPs represents the genetics underlying kidney function.

Methods. In the community-based PREVEND Study, we assessed $eGFR_{crea}$ and $eGFR_{cysc}$ at baseline and four follow-up examinations. The GRS comprised 53 SNPs for $eGFR_{crea}$ weighted for reported effect-sizes. We adjusted for baseline demographics and renal risk factors.

Results. We included 3649 subjects (median age 49 years, 52% male, median follow-up 11 years, N=85 baseline CKD, N=154 incident CKD). At baseline, a higher GRS associated with lower $eGFR_{crea}$ (adjusted B (95%CI) = -2.05 (-2.45;-1.65) mL/min/1.73m^2, p<0.001) and higher CKD prevalence (adjusted OR (95%CI)= 1.41 (1.12;1.77), p=0.002). During follow-up, a higher GRS associated with higher CKD incidence (adjusted HR (95%CI)= 1.28 (1.09;1.50), p=0.004), but no longer significantly after adjustment for baseline eGFR. No significant association with $eGFR_{crea}$ decline was found. Associations with cystatin C-estimated outcomes were similar.

Conclusions. The GRS robustly associated with baseline CKD and eGFR, independent of known risk factors. Associations with incident CKD were likely due to low baseline eGFR, not accelerated eGFR decline. The GRS for $eGFR_{crea}$ likely represents the genetics underlying kidney function, not creatinine metabolism or underlying etiologies. To improve clinical utility of GWAS results for CKD, these need to specifically address eGFR decline and CKD incidence.

INTRODUCTION

Chronic kidney disease (CKD) is a heterogeneous group of diseases defined by the presence of sustained reduced kidney function or kidney damage. Strong evidence exists for a genetic component to CKD risk: CKD has been observed to aggregate in families[1-3] and heritability estimates are reported to range between 30 and 75%[4-9]. Furthermore, genome-wide association studies (GWAS) in populations of European ancestry have identified common genetic variants associated with CKD and kidney function markers[10-14] . The largest and most comprehensive genetic study is a cross-sectional meta-analysis of GWASs, in which single nucleotide polymorphisms (SNPs) at 53 loci were found to be associated with creatinine-estimated eGFR ($eGFR_{crea}$)[15].

The individual SNPs identified in this meta-analysis can be combined into a genetic risk score (GRS)[16-18], which summarizes individual genetic predisposition to CKD. Such a GRS is a potentially useful tool in etiological and predictive studies of CKD. However, because the SNPs were identified in a cross-sectional GWAS design, it is uncertain whether a GRS is associated with longitudinal outcomes. Furthermore, there is overlap between the 53 loci from the aforementioned meta-analysis and loci identified in a large GWAS on serum creatinine[11,12]. Therefore, it is difficult to discern whether a GRS corresponds to kidney function per se or partly reflects creatinine production/secretion.

The main study aim was to evaluate the applicability of a GRS, comprising 53 SNPs identified in cross-sectional GWAS on $eGFR_{crea}$, in longitudinal outcomes. To this end, we tested three hypotheses. First, we tested the hypothesis that the GRS would be associated with kidney outcomes, not only cross-sectionally (i.e. with baseline CKD, baseline eGFR), but also longitudinally (i.e. with incident CKD, eGFR decline). Second, to assess whether the GRS is a true representation of a genetic component to kidney function, we hypothesized that the GRS would also be associated with GFR estimates not based on serum creatinine. We therefore compared the associations of the GRS with $eGFR_{crea}$ to those of the GRS with an serum cystatin C-estimated GFR ($eGFR_{cysc}$)[19]. Third, to rule out that the GRS represents a component to kidney damage rather than kidney function, we hypothesized that the GRS would not be associated with albuminuria (i.e. urinary albumin excretion, UAE).

METHODS

Study population and design

We used data from the Prevention of REnal and Vascular ENdstage Disease (PREVEND) cohort study[20]. PREVEND was initiated to investigate the natural course of increased urinary albumin levels and its association to renal and vascular outcomes. Details of this study have been described elsewhere. In brief, 8592 individuals, sampled from the general population of Groningen, the Netherlands, underwent extensive examination between 1997-1998. The four follow-up examinations were completed in 2003, 2006, 2008, and 2012. Included were 3649 subjects of whom GWAS data were available. All subjects gave written informed consent. The PREVEND Study was approved by the medical ethics committee of the University Medical Center Groningen and conducted in accordance with the Helsinki Declaration guidelines.

Genetic risk scores

Genotyping details for PREVEND were described previously[21]. In brief, genotyping was performed on the Illumina CytoSNP12 v2 chip. Variants were imputed to 1000G[22], phase 1 version 3, using Minimac software[23]. Population stratification was assessed by principal component analysis; samples with Z-score>3 for any of the first five principal components were excluded, i.e. outlying individuals were removed because of likely divergent ancestry[24]. Samples with a call rate<95%, duplicates, and sex discrepancies were excluded. Markers with call rate>95%, Hardy-Weinberg equilibrium p-value≥$1x10^{-5}$, and minor allele frequency (MAF)≥1% were included. From the resulting GWAS data, we extracted the genotypes of the 53 SNPs that were identified in a recent meta-analysis of GWAS on $eGFR_{crea}$ in European populations[15]. Designated risk alleles were those associated with lower eGFR. Genotypes were represented as continuous allelic dosages from 0 to 2, reflecting an additive model [25]. A weighted GRS was defined as the sum of the risk alleles weighted for their published regression coefficient. Therefore, a higher GRS corresponds to higher susceptibility to impaired kidney function. For ease of interpretation, effects are reported per standard deviation (sd) higher GRS.

Outcome measurements and definition

At each examination, participants collected two consecutive 24h-urine specimens after thorough instruction. Participants were asked to avoid heavy exercise as

much as possible before urine collection, and instructed to postpone urine collection in case of urinary tract infection, menstruation, or fever. The collected urine was stored cold (4°C) for a maximum of four days before handing it in. After this, urine specimens were stored at -20°C. Fasting blood samples were obtained and stored at -80°C.

Measurement of serum creatinine was performed by an enzymatic method on a Roche Modular analyzer using reagents and calibrators from Roche (Roche Diagnostics, Mannheim, Germany), with intra- and interassay coefficients of variation of 0.9% and 2.9%, respectively. Serum cystatin C concentration was measured by a Gentian cystatin C Immunoassay (Gentian AS Moss, Norway) on a Modular analyzer (Roche Diagnostics). Cystatin C was calibrated directly using the standard supplied by the manufacturer (traceable to the International Federation of Clinical Chemistry Working Group for Standardization of Serum Cystatin C)[26] The intra- and interassay coefficients of variation were <4.1% and <3.3%, respectively. Urinary albumin concentration (UAC) was measured by nephelometry with a lower threshold of detection of 2.3mg/L, and intra- and interassay coefficient of variation of 2.2% and 2.6%, respectively (Dade Behring Diagnostic, Marburg, Germany). UAC was multiplied by urine volume to obtain a value of UAE in mg/24h. The two 24h-urinary albumin values of each subject per examination were averaged.

We calculated $eGFR_{crea}$ from serum creatinine and $eGFR_{cysc}$ from serum cystatin C, using the corresponding CKD-EPI equations [19]. We defined CKD_{crea} as $eGFR_{crea}$<60ml/min/1.73m^2, CKD_{cysc} as $eGFR_{cysc}$<60ml/min/1.73m^2, and CKD_{UAE} as UAE≥30mg/24h. Incident cases were those free of CKD at baseline who developed CKD during follow-up. In secondary analyses, we used the CKD-EPI equation for both serum creatinine and cystatin C to calculate $eGFR_{crea\text{-}cysc}$ [27]. Furthermore, a definition of CKD based on KDIGO guidelines (CKD_{KDIGO}, $eGFR_{crea\text{-}cysc}$<60ml/min/1.73m^2 and/or UAE≥30mg/24h) was used[28].

Covariates

We selected the following renal risk factors as covariates: age, sex, body-mass index (BMI, weight/height2 [kg/m^2]), current smoking (self-reported yes/no), diabetes (fasting glucose>7.0mmol/L, non-fasting glucose>11.0mmol/L, anti-diabetic treatment, or self-reported), hypertension (systolic blood pressure>140mmHg, diastolic blood pressure>90mmHg, blood pressure lowering treatment, or self-

reported), hypercholesterolemia (total cholesterol≥6.21mmol/L, lipid lowering treatment, or self-reported), and history of cardiovascular disease (CVD, any past cardio/cerebrovascular event or intervention). Covariates were collected at baseline by means of questionnaires, anthropometry, and pharmacy records.

Statistical analyses

Analyses were performed using R3.3.1 and SPSS23.0 (IBM Corporation). Two-sided significance level for analyses was set at α=0.05 unless stated otherwise.

Baseline characteristics

Baseline characteristics were examined for the total population. One-way ANOVA, Jonckheere-Terpstra, and χ^2-tests were used to examine linear trends of characteristics across tertiles of GRS. In subsequent analyses, GRS was treated as a continuous variable. We examined age and sex-adjusted associations of all 53 individual SNPs with baseline $eGFR_{crea}$ and $eGFR_{cysc}$ using ordinary least squares (OLS) regression.

Cross-sectional associations of the GRS with CKD prevalence and baseline eGFR

Logistic regression was used to examine the association of the continuous GRS with baseline CKD_{crea}. We adjusted for covariates by adding incremental groups of covariates in order to distinguish confounding effects of demographics and risk factors. Group 1 consisted of age and sex; group 2 additionally included BMI, smoking, diabetes, hypertension, hypercholesterolemia, and history of CVD.

We examined the association of the GRS with continuous $eGFR_{crea}$ using OLS regression. We adjusted for covariates as described above. Analyses were repeated for baseline $eGFR_{cysc}$ and prevalent CKD_{cysc}.

Longitudinal associations of the GRS with CKD incidence and eGFR decline

Cox regression models were used to examine the association of continuous GRS with incident CKD_{crea}. To estimate time to incident CKD_{crea}, we used a midpoint imputation technique. In this analysis, we corrected for baseline $eGFR_{crea}$ in addition to the previously listed renal risk factors. Subjects were censored at death or date of last visit.

Linear mixed-effects (LME) analysis was performed to examine the association of the GRS with eGFR decline. We modelled $eGFR_{crea}$ as a function of time since

baseline (per year). We specified a model with random intercept, random coefficient for time, and unstructured covariance matrix. The GRS, time, and covariates were included as fixed effects. A two-way interaction term between GRS and time was introduced to assess whether $eGFR_{crea}$ decline differed by values of the GRS. Analyses were repeated with the outcomes $eGFR_{cysc}$ decline and incident CKD_{cysc}.

Associations with UAE

We repeated the cross-sectional and longitudinal analyses described above to examine associations of a GRS with renal outcomes based on elevated UAE. Continuous UAE was transformed by its natural logarithm to approach normality (ln(UAE)), in OLS regression and LME analyses.

Secondary analyses

We repeated all analyses using $eGFR_{crea\text{-}cysc}$ and CKD_{KDIGO} as outcome. Furthermore, we constructed two alternative GRS. The first alternative GRS comprised 49 SNPs that were significant in the meta-analysis by Gorski et al.[14], with the second alternative comprising all 63 SNPs identified in either the Pattaro (53 SNPs) and the Gorski study (10 additional SNPs).

RESULTS

Baseline characteristics

Baseline characteristics of the 3649 subjects are presented in **Table 1**. In univariable analyses, a higher tertile for the GRS was associated with higher serum creatinine and cystatin C levels (p_{trend}<0.001); higher prevalence of CKD_{crea} (p_{trend} =0.002) and CKD_{cysc} (p_{trend} =0.01); lower $eGFR_{crea}$ (p_{trend} end<0.001) and lower $eGFR_{cysc}$ (p_{trend} <0.001); lower UAE (p_{trend} <0.001). No associations with CKD_{UAE} were found. We found no associations with age, sex, BMI, smoking status, diabetes, hypertension, hypercholesterolemia, or history of CVD.

Details of the 53 SNPs used in the calculation of the GRS and age- and sex- adjusted estimates of their association to baseline $eGFR_{crea}$, baseline $eGFR_{cysc}$, and ln(UAE) are listed in **Supplementary Table S1A**. Out of 53 SNPs, 22 reached nominal significance (one-sided p<0.05), while three were significant when a Bonferroni correction for 53 tests ($p<9.4x10^{-4}$) was applied.

Table 1. Baseline characteristics of the cohort stratified by tertiles of the Genetic Risk Score

	Total	GRS			P_{trend}
		low	medium	high	
N	3649	1216	1217	1216	n/a
Age, years	49 [39-60]	49 [40-60]	49 [39-59]	49 [39-60]	0.954
Males, %	52%	51%	51%	52%	0.598
BMI, kg/m²	26 (4.3)	26 (4.2)	26 (4.4)	26 (4.2)	0.816
BMI ≥30, %	16%	16%	16%	16%	0.868
Current smoker, %	35%	35%	35%	36%	0.420
Hypertension, %	34%	32%	36%	34%	0.521
SBP, mmHg	129 (20)	129 (20)	129 (20)	129 (20)	0.612
DBP, mmHg	74 (9.9)	74 (9.9)	74 (10)	74 (9.9)	0.887
BP lowering medication, %	12%	13%	14%	11%	0.658
Diabetes, %	3.9%	3.7%	3.4%	4.7%	0.210
Glucose, mmol/L	4.7 [4.3-5.1]	4.7 [4.4-5.1]	4.7 [4.3-5.2]	4.7 [4.4-5.1]	0.926
Anti-diabetic medication, %	1.3%	1.2%	1.0%	1.8%	0.843
Hypercholesterolemia, %	31%	31%	32%	31%	1.000
Total cholesterol, mmol/L	5.7 (1.1)	5.7 (1.1)	5.6 (1.1)	5.7 (1.1)	0.744
Lipid lowering medication, %	3.6%	4.8%	3.7%	2.7%	0.499
History of CVD, %	4.2%	3.9%	5.1%	3.8%	0.920
Serum creatinine, mg/dL	0.82 (0.18)	0.79 (0.16)	0.82 (0.18)	0.85 (0.19)	<0.001
$eGFR_{crea}$, mL/min/1.73m²	96 (16)	98 (15)	96 (16)	94 (16)	<0.001
CKD_{crea}: $eGFR_{crea}$ <60, %	2.5%	1.4%	2.7%	3.4%	0.002
Serum cystatin C, mg/L	0.90 (0.18)	0.88 (0.17)	0.90 (0.19)	0.92 (0.18)	<0.001
$eGFR_{cysc}$, mL/min/1.73m²	92 (19)	94 (19)	92 (19)	90 (19)	<0.001
CKD_{cysc}: $eGFR_{cysc}$ <60, %	5.9%	4.9%	5.3%	7.4%	0.010
$eGFR_{crea\text{-}cysc}$	94 (17)	97 (17)	95 (17)	92 (17)	<0.001
CKD_{KDIGO} :$eGFR_{crea\text{-}cysc}$ <60 or UAE≥30 , %	20%	21%	20%	19%	0.297
UAE, mg/24h	10.6 [6.6-21]	11.5 [7.0-23]	10.3 [6.6-20]	10.2 [6.4-20]	<0.001
UAE ≥30 , %	17%	19%	17%	17%	0.172
No of risk alleles	57 (4.5)	52 (2.7)	57 (1.7)	62 (2.5)	<0.001

Baseline characteristics of the cohort. Data is presented as mean (standard deviation), median [interquartile range], and percentage where appropriate. P-values for linear trend were calculated using one-way ANOVA, Jonckheere-Terpstra-tests, and χ^2-tests where appropriate.

Abbreviations: BMI, body mass index; CVD, cardiovascular disease; CKD, chronic kidney disease; DBP, diastolic blood pressure; eGFR, estimated glomerular filtration rate; GRS, genetic risk score; SBP, systolic blood pressure; UAE, urinary albumin excretion

Supplementary Figure S2 presents a plot of age- and sex-adjusted regression coefficients. These coefficients were obtained by OLS regression of individual SNPs on either $eGFR_{crea}$ and $eGFR_{cysc}$. Correlation between the regression coefficients on $eGFR_{crea}$ and $eGFR_{cysc}$ was moderate (Pearson r=0.51, p<0.001). The total least squares regression line showed fair agreement with the line of identity.

Cross-sectional associations of the GRS with baseline eGFR and CKD prevalence

We present cross-sectional results in **Table 2**. Per sd higher GRS, the odds of having CKD_{crea} at baseline increased by 41% (fully adjusted odds ratio (OR) (95%CI)=1.41 (1.12;1.77), p=0.002. A higher GRS was associated with lower $eGFR_{crea}$ (fully adjusted unstandardized coefficient B (95%CI)= -2.05 (-2.45;-1.65) mL/min/1.73m², p<0.001), independent of known risk factors. Effect sizes of the associations with CKD_{cysc} (adjusted OR (95%CI)= 1.27 (1.08;1.50), p=0.004) and with $eGFR_{cysc}$ (adjusted B (95%CI)= -1.63 (-2.11;-1.14) mL/min/1.73m², p<0.001) were smaller but showed a similar trend compared to those for creatinine-estimated outcomes. Estimates of the effect sizes of the GRS on both $eGFR_{crea}$ and $eGFR_{cysc}$ remained stable during incremental covariate adjustment.

Longitudinal associations of the GRS with eGFR decline and CKD incidence

We present longitudinal results in **Table 3**. A higher GRS was associated with higher incidence of CKD_{crea} after adjustment for known renal risk factors (adjusted hazard ratio (HR) (95%CI)=1.28 (1.09;1.50), p=0.003), but significance disappeared after additional adjustment for baseline $eGFR_{crea}$ (fully adjusted HR (95%CI)=1.05 (0.89;1.24), p=0.537). A higher GRS was not associated with steeper decline of $eGFR_{crea}$ (fully adjusted B (95%CI)= -0.01 (-0.04;0.03) mL/min/1.73m² per year, p=0.655). Inclusion of interaction terms between baseline renal risk factors and time did not change estimates of the effects between the GRS and eGFR decline (data not shown).

Similar associations were found with $eGFR_{cysc}$ decline (fully adjusted B (95%CI)= -0.03 (-0.07;0.01) mL/min/1.73m² per year, p=0.167) and incident CKD_{cysc} (adjusted HR (95%CI)=1.17 (1.03;1.32), p=0.014. The association with incident CKD_{cysc} lost significance after additional adjustment for baseline $eGFR_{cysc}$ (fully adjusted HR (95%CI)=1.06 (0.94;1.20), p=0.336).

Table 2. Cross-sectional associations of the Genetic Risk Score with selected kidney outcomes at baseline

	Dichotomous outcomes					Continuous outcomes			
	Prevalent CKD_{crea} (85 cases / N= 3397)	**Prevalent CKD_{cysc}** (199 cases / N=3394)	**Prevalent CKD_{UAE}** (635 cases / N=3614)	**Prevalent CKD_{KDIGO}** (684 cases / N=3423)		**$eGFR_{crea}$** (N=3397)	**$eGFR_{cysc}$** (N=3394)	**$eGFR_{crea-cysc}$** (N=3394)	**ln(UAE)** (N=3614)
	OR (95%CI)	**OR (95%CI)**	**OR (95%CI)**	**OR (95%CI)**		**B (95%CI)**	**B (95%CI)**	**B (95%CI)**	**B (95%CI)**
Model 1	1.38 (1.11;1.71)**	1.21 (1.05;1.40)**	0.92 (0.85;1.00)	0.93 (0.86;1.01)		-1.87 (-2.39; -1.34)***	-1.45 (-2.10;-0.81)***	-1.85 (-2.42;-1.28)***	-0.05 (-0.08;-0.01)**
Model 2	1.41 (1.13;1.76)**	1.27 (1.08;1.48)**	0.93 (0.85;1.01)	0.94 (0.86;1.02)		-2.04 (-2.44;-1.64)***	-1.66 (-2.15;-1.16)***	-2.04 (-2.46;-1.61)***	-0.04 (-0.07;-0.01)**
Model 3	1.41 (1.12;1.77)**	1.27 (1.08;1.50)**	0.92 (0.84;1.01)	0.93 (0.85;1.02)		-2.05 (-2.45;-1.65)***	-1.63 (-2.11;-1.14)***	-2.02 (-2.45;-1.60)***	-0.04 (-0.07;-0.01)**

Estimates from linear and logistic regression analyses. Data is presented as regression coefficient B (95% confidence interval), or odds ratio OR (95% confidence interval), per standard deviation (sd) of GRS. Definitions and abbreviations: eGFR, estimated glomerular filtration rate (mL/min/1.73m²) ; ln(UAE), natural logarithm (ln) of urinary albumin excretion (ln mg/24h) ; $CKD_{crea/cysc}$, chronic kidney disease ($eGFR_{crea/cysc}$ <60mL/min/1.73m²) ; CKD_{UAE} (UAE ≥30 mg/24h) ; CKD_{KDIGO} , ($eGFR_{crea-cysc}$ <60mL/min/1.73m² and/or UAE ≥30 mg/24h) ; GRS, genetic risk score. *p<0.05, **p<0.01, ***p<0.001.

Model 1: GRS
Model 2: model 1 + age + sex,
Model 3: model 2 + BMI + smoking + diabetes + hypertension + hypercholesterolemia + history of cardiovascular disease.

Table 3. Longitudinal associations of the Genetic Risk Score with selected kidney outcomes during follow-up

	Dichotomous outcomes					Continuous outcomes			
	Incident CKD_{crea} (154 cases / N= 2731)	**Incident CKD_{cysc}** (279 cases / N=2659)	**Incident CKD_{UAE}** (368 cases / N=2493)	**Incident CKD_{KDIGO}** (411 cases / N=2296)		**Δ$eGFR_{crea}$** (N=3447)	**Δ$eGFR_{cysc}$** (N=3447)	**Δ$eGFR_{crea-cysc}$** (N=3447)	**Δln(UAE)** (N=3619)
	HR (95%CI)	**HR (95%CI)**	**HR (95%CI)**	**HR (95%CI)**		**B (95%CI)**	**B (95%CI)**	**B (95%CI)**	**B (95%CI)**
Model 1	1.19 (1.02;1.40)*	1.08 (0.96;1.21)	0.94 (0.85;1.04)	1.02 (0.93;1.13)		-0.01 (-0.04;0.03)	-0.03 (-0.07;0.01)	-0.02 (-0.05;0.02)	0.001 (-0.001;0.004)
Model 2	1.28 (1.09;1.50)**	1.15 (1.02;1.29)*	0.96 (0.87;1.06)	1.07 (0.97;1.18)		-0.01 (-0.04;0.03)	-0.03 (-0.07;0.01)	-0.02 (-0.05;0.02)	0.001 (-0.001;0.004)
Model 3	1.28 (1.09;1.50)**	1.17 (1.03;1.32)*	0.95 (0.86;1.06)	1.06 (0.96;1.17)		-0.01 (-0.04;0.03)	-0.03 (-0.07;0.01)	-0.02 (-0.05;0.02)	0.001 (-0.001;0.004)
Model 4	1.05 (0.89;1.24)	1.06 (0.94;1.20)	1.03 (0.93;1.14)	1.11 (1.00;1.22)^		-	-	-	-

Estimates from Cox regression and LME analyses. Data is presented as hazard ratio HR (95% confidence interval), or regression coefficient B (95% confidence interval), per standard deviation of GRS. Definitions and abbreviations: Δ eGFR, annual change in estimated glomerular filtration rate (mL/min/1.73m² per year) ; Δ ln(UAE), annual change in natural logarithm (ln) of urinary albumin excretion (ln (mg/24h) per year) ; $CKD_{crea/cysc}$, chronic kidney disease ($eGFR_{crea/cysc}$ <60mL/min/1.73m²) ; CKD_{UAE} (UAE ≥30 mg/24h) ; CKD_{KDIGO} , ($eGFR_{crea-cysc}$ <60mL/min/1.73m² and/or UAE ≥30 mg/24h); GRS, genetic risk score. *p<0.05, **p<0.01, ***p<0.001.

Model 1: GRS
Model 2: model 1 + age + sex
Model 3: model 2 + BMI + smoking + diabetes + hypertension + hypercholesterolemia + history of cardiovascular disease
Model 4: model 3 + baseline eGFR (adjusted for baseline UAE instead of eGFR) (^adjusted for both baseline eGFR and UAE)

Association of the GRS with UAE

Results of analyses on UAE are presented in **Table 2-3**. A higher GRS was associated with lower ln(UAE) (fully adjusted B (95%CI)= -0.04 (-0.07;-0.01) ln(mg/24h), p=0.004) but not with higher prevalence of CKD_{UAE} (fully adjusted OR (95%CI)=0.92 (0.84;1.01), p=0.074). No longitudinal associations of GRS with kidney damage were observed: a higher GRS was neither associated with steeper increase of ln(UAE) (fully adjusted B (95%CI)=0.001 (-0.001;0.004) ln(mg/24h) per year, p=0.297) nor with higher incidence of CKD_{UAE} (fully adjusted HR (95%CI)=1.03 (0.93;1.14), p=0.360).

Analyses with 24h-urinary albumin-to-creatinine ratio as outcome yielded similar results (data not shown).

Secondary analyses

Associations of the GRS with $eGFR_{crea-cysc}$ were consistent with those of the GRS with $eGFR_{crea}$ and $eGFR_{cysc}$. We found no cross-sectional or longitudinal association of the GRS with CKD_{KDIGO} (Table 2-3). Two alternative GRS, based on 49 SNPs ($GRS_{1000G-49}$) and 63 SNPs ($GRS_{1000G-63}$), were evaluated. Individual SNP-effects of these GRS are listed in **Supplementary Table S1B**. The GRSs showed similar but slightly weaker associations compared to our main GRS (**Supplementary Table S3-7**).

DISCUSSION

In this population based, longitudinal cohort study, we evaluated the effects of a GRS comprising 53 $eGFR_{crea}$-SNPs on kidney outcomes. To this end, we tested cross-sectional and longitudinal associations of this GRS with CKD_{crea} and $eGFR_{crea}$ and compared these associations to those with CKD_{cysc} and $eGFR_{cysc}$. Cross-sectional associations of the GRS with the kidney outcomes, CKD_{crea} and $eGFR_{crea}$, were modest but robust, corroborating the literature. In longitudinal analyses, we observed no associations with kidney function decline. The GRS was associated with incidence of CKD_{crea}, but this was likely due to lower baseline eGFR rather than accelerated kidney function decline. In comparison to associations with $eGFR_{crea}$, associations with $eGFR_{cysc}$ were smaller but showed a similar trend. Higher GRS was not associated with kidney damage markers. Furthermore, all associations of the GRS with kidney outcomes were independent of renal risk factors. These data suggest that the GRS is a true representation of the genetics underlying kidney function, as opposed to creatinine metabolism, kidney damage, or related etiologies such as hypertension/diabetes.

In secondary analyses, we confirmed associations with $eGFR_{crea-cysc}$, currently the best estimate for kidney function for large population-based studies [19,29]. We found no association of the GRS with CKD_{KDIGO} as outcome. This is likely due to the fact that this GRS was optimized for eGFR as outcome and not urinary albumin; in our sample, CKD_{KDIGO} was predominantly characterized by elevated urinary albumin rather than diminished kidney function. Two alternative GRS ($GRS_{1000G-49}$ and $GRS_{1000G-63}$), yielded similar results but proved to be slightly less powerful predictors of kidney function and CKD in this sample.

Previously, two similar GRSs based on $eGFR_{crea}$ SNPs were investigated in ~2500 participants with ~11 years of follow-up from the Framingham Heart Study. O'Seaghdha et al. calculated a 16-SNP GRS for $eGFR_{crea}$ [17]. This sample of the Framingham cohort was revisited by Ma et al.[18], who updated the GRS with 37 additional SNPs, that is the same 53 as the present study. Both of these GRS were independently associated with incident CKD ($eGFR_{crea}$<60mL/min/1.73m^2), although neither of these GRSs improved prediction and/or discrimination beyond clinical risk factors (age, sex, BMI, eGFR, hypertension, diabetes, proteinuria). Interestingly, they reported associations of a higher GRS with a higher incidence of CKD to be independent of baseline eGFR, hence an accelerated deterioration of kidney function in those with a higher GRS. Such an effect was also suggested by Böger et al.[30] in a study of eGFR related loci identified by GWAS. In 26,308 individuals of European ancestry, the associations of 16 separate SNPs known at the time with incident CKD were examined. Of these 16 SNPs, six (mapping to *UMOD*, *PRKAG2*, *LASS2*, *DAB2*, *DACH1*, and *STC1*) were significantly ($p<0.05$) associated with incident CKD (eGFR<60mL/min/1.73m^2), even after correction for baseline eGFR. Similar to the findings of O'Seaghdha and Ma et al, this implies that several SNPs associate with eGFR decline. In contrast, in the present study we could not corroborate such an effect on CKD incidence or eGFR decline: the association of GRS with incident CKD was not significant after adjustment for baseline eGFR, and there was no significant association between the GRS and eGFR decline.

A possible explanation for this discrepancy is the potential overestimation of the effect of the GRS by O'Seaghdha and Ma et al. due to the participation of the Framingham Cohort Study in the discovery phase of the meta-analysis[12,15]. Similarly, overestimation of individual SNP effects may have occurred in the

study by Böger et al, given that seven of the eight cohorts participating in that study were part of the discovery GWAS[12]. Such overlap in discovery and validation cohorts might result in inflated effect sizes[31]. The PREVEND study was not part of the original discovery GWAS, ensuring its independence and suitability as a validation cohort for evaluation of a GRS based on $eGFR_{crea}$ SNPs. This potential overestimation possibly also explains that in our study, the GRS explained only 1.66% of variance of baseline $eGFR_{crea}$, whereas in the original GWAS, the explained variance of $eGFR_{crea}$ by the combined loci was 3.22%[15].

Notwithstanding these discrepancies, the combined data suggest that the genetics underlying kidney function are, at least partly, distinct from that underlying kidney function decline and/or kidney disease susceptibility. Our results indicate that a GRS based on cross-sectional GWAS results on kidney function is not clinically applicable (e.g. in the prediction of CKD risk). A GRS would be more applicable if SNPs associated with kidney function decline and/or CKD incidence were used, as these would likely better represent disease susceptibility. Unfortunately, there is paucity of data on genetic loci associated with kidney function decline or CKD incidence. To the best of our knowledge, only one study by Gorski et al. performed a GWAS for kidney function decline phenotypes[32]. In this study, only one SNP mapping to *UMOD* (which was also implicated in prior GWAS on cross-sectional $eGFR_{crea}$) was significantly associated with eGFR change in the general population, while two novel loci, *CDH23* and *GALNT15/GALNT11* were only suggestively associated with eGFR change in CKD patients, and rapid decline in the general population, respectively. To benefit clinical applicability, we argue that future GWAS should focus on disease susceptibility genes, i.e. loci associated with eGFR decline and/or CKD incidence. We found a higher GRS to be associated with lower UAE, i.e. lower risk of kidney damage, which is surprising for two reasons. First, a prior family study, using bivariate variance component linkage analysis techniques, found a low genetic correlation between eGFR and UACR (r_g=0.002 in African Americans, not reported for European Americans)[5]. Second, there is no overlap in genome-wide significant markers for eGFR and albuminuria in the general population[33, 34]. Due to this apparent lack of genetic overlap, it is believed that eGFR and albuminuria have distinct genetic underpinnings. To our knowledge, we are the first to observe this counterintuitive association with the updated 53 SNP GRS. Although the correlation between the GRS and ln(UAE) was weak (r=-0.043), it is unlikely to be a chance finding: in an earlier study by Ellis et al.

a weighted GRS (comprising 16 SNPs associated with eGFRcrea) was associated with both lower eGFR and with lower UACR[35]. The authors attributed this effect to the A-allele of rs17319721, a SNP mapping to *SHROOM3*, because exclusion of this SNP from their GRS attenuated the effect on UACR. In previous GWAS, the *SHROOM3* SNP was found to be associated with $eGFR_{crea}$[12], and suggestively with UACR ($p=7.0x10^{-7}$)[15,33,34]. In the present study, exclusion of this SNP from the updated GRS did not attenuate the effect (data not shown). Therefore, it is possible that, in addition to *SHROOM3*, other loci discovered in the recent meta-analysis on $eGFR_{crea}$ might have pleiotropic effects on both eGFR and albuminuria. We therefore performed a query in LDHub v1.3.1, a platform for LD-score regression which uses original GWAS summary statistics[36, 37]. LD Hub showed a modest genetic correlation between $eGFR_{crea}$ and UACR ($r_g=0.388$, $p<0.001$), and a suggestive genetic correlation between $eGFR_{cysc}$ and UACR ($r_g=0.195$, $p=0.087$), in the same direction as our findings (i.e. higher eGFR~higher UACR). These correlations suggest that there is at least partial overlap in the genetics underlying eGFR and albuminuria. Addressing the question of pleiotropy is beyond the scope of the present study and requires dedicated analysis in larger samples.

A number of SNPs identified in the GWAS on $eGFR_{crea}$ may be linked to loci related to creatinine production or secretion, hence not with kidney function per se[38]. We therefore examined two SNPs mapping to loci known to be related to creatinine metabolism: rs2467853 which maps to the creatinine production locus *GATM*[39] and rs316009 which maps to the creatinine secretion locus SLC22A2[40]. For both SNPs, we observed an inconsistency in the direction of effect for baseline $eGFR_{crea}$ and $eGFR_{cysc}$ (see **Supplementary Table 1A**), suggesting that these loci are indeed not related to kidney function. Exclusion of these SNPs led to a slightly improved GRS: effects of this GRS on $eGFR_{crea}$ and $eGFR_{cysc}$ more closely resembled each other than those of the main GRS, although this improvement was only slight (data not shown). Our conclusions therefore remain unchanged. Future, functional studies may investigate other presumptive creatinine-related loci. The exclusion of such loci may result in a GRS that more accurately reflects genetic predisposition to kidney function.

To our knowledge, we are the first study that examined the association between a GRS comprising 53 SNPs and eGFR decline. Strengths of this study include the availability of serially measured creatinine and cystatin C, as well as two 24h-urinary

albumin at each examination, during a considerable follow-up duration of 11 years. A major strength of PREVEND is its independence from the discovery GWAS that identified the 53 SNPs used in the GRS, resulting in unbiased effect estimates of the GRS. Given that participants of the PREVEND GWAS sample are of European ancestry, we cannot generalize to other ethnicities. Finally, we could not calculate genetic correlations between eGFR levels and eGFR decline as GWAS summary results for eGFR decline were currently not available.

In conclusion, a GRS comprising 53 SNPs showed modest but robust associations with cross-sectional CKD outcomes based on $eGFR_{crea}$. These associations were confirmed with $eGFR_{cysc}$, which highlights the potential usefulness of a GRS as a representation of the genetics underlying kidney function. However, no longitudinal associations with incident CKD or eGFR decline were found. Given these results, we question the clinical utility of cross-sectional GWAS results on kidney function. We suggest that future GWAS specifically examine genetic associations with eGFR decline and/or CKD incidence. These GWAS may identify loci that, when incorporated into a GRS, will improve the clinical utility of this score, e.g. in predicting onset of CKD.

CONFLICT OF INTEREST STATEMENT

None of the authors declare a conflict of interests.

ACKNOWLEDGEMENTS

The PREVEND Study was funded by grants from the Dutch Kidney Foundation

All Supplementary material can be accessed via the following link:
www.academic.oup.com/ndt/article/33/10/1757/4774601

REFERENCES

1. Skrunes R, Svarstad E, Reisaeter AV, Vikse BE. Familial clustering of ESRD in the norwegian population. *Clin J Am Soc Nephrol*. 2014;9(10):1692-1700.

2. Satko SG, Sedor JR, Iyengar SK, Freedman BI. Familial clustering of chronic kidney disease. . 2007;20(3):229-236.

3. Lei HH, Perneger TV, Klag MJ, Whelton PK, Coresh J. Familial aggregation of renal disease in a population-based case-control study. *J Am Soc Nephrol*. 1998;9(7):1270-1276.

4. Arpegard J, Viktorin A, Chang Z, de Faire U, Magnusson PK, Svensson P. Comparison of heritability of cystatin C- and creatinine-based estimates of kidney function and their relation to heritability of cardiovascular disease. *J Am Heart Assoc*. 2015;4(1):e001467.

5. Leon JM, Freedman BI, Miller MB, et al. Genome scan of glomerular filtration rate and albuminuria: The HyperGEN study. *Nephrol Dial Transplant*. 2007;22(3):763-771.

6. Fox CS, Yang Q, Cupples LA, et al. Genomewide linkage analysis to serum creatinine, GFR, and creatinine clearance in a community-based population: The framingham heart study. *J Am Soc Nephrol*. 2004;15(9):2457-2461.

7. Hunter DJ, Lange M, Snieder H, et al. Genetic contribution to renal function and electrolyte balance: A twin study. *Clin Sci (Lond)*. 2002;103(3):259-265.

8. Langefeld CD, Beck SR, Bowden DW, Rich SS, Wagenknecht LE, Freedman BI. Heritability of GFR and albuminuria in caucasians with type 2 diabetes mellitus. *Am J Kidney Dis*. 2004;43(5):796-800.

9. Bochud M, Elston RC, Maillard M, et al. Heritability of renal function in hypertensive families of african descent in the seychelles (indian ocean). *Kidney Int*. 2005;67(1):61-69.

10. Kottgen A, Glazer NL, Dehghan A, et al. Multiple loci associated with indices of renal function and chronic kidney disease. *Nat Genet*. 2009;41(6):712-717.

11. Chambers JC, Zhang W, Lord GM, et al. Genetic loci influencing kidney function and chronic kidney disease. *Nat Genet*. 2010;42(5):373-375.

12. Kottgen A, Pattaro C, Boger CA, et al. New loci associated with kidney function and chronic kidney disease. *Nat Genet*. 2010;42(5):376-384.

13. Gudbjartsson DF, Holm H, Indridason OS, et al. Association of variants at UMOD with chronic kidney disease and kidney stones—role of age and comorbid diseases. *PLoS Genet*. 2010;6(7):e1001039.

14. Gorski M, van der Most PJ, Teumer A, et al. 1000 genomes-based meta-analysis identifies 10 novel loci for kidney function. *Scientific Reports*. 2017;7:45040.

15. Pattaro C, Teumer A, Gorski M, et al. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat Commun*. 2016;7.

16. Morrison AC, Bare LA, Chambless LE, et al. Prediction of coronary heart disease risk using a genetic risk score: The atherosclerosis risk in communities study. *Am J Epidemiol*. 2007;166(1):28-35.

17. O'Seaghdha CM, Yang Q, Wu H, Hwang SJ, Fox CS. Performance of a genetic risk score for CKD stage 3 in the general population. *Am J Kidney Dis*. 2012;59(1):19-24.

18. Ma J, Yang Q, Hwang S, Fox CS, Chu AY. Genetic risk score and risk of stage 3 chronic kidney disease. *BMC nephrology*. 2017;18(1):32.

19. Inker LA, Schmid CH, Tighiouart H, et al. Estimating glomerular filtration rate from serum creatinine and cystatin C. *N Engl J Med*. 2012;367(1):20-29.

20. Pinto-Sietsma SJ, Janssen WM, Hillege HL, Navis G, De Zeeuw D, De Jong PE. Urinary albumin excretion is associated with renal functional abnormalities in a nondiabetic population. *J Am Soc Nephrol*. 2000;11(10):1882-1888.

21. Verweij N, Mateo Leach I, Isaacs A, et al. Twenty-eight genetic loci associated with ST-T-wave amplitudes of the electrocardiogram. *Hum Mol Genet*. 2016;25(10):2093-2103.

22. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467(7319):1061-1073.

23. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet*. 2012;44(8):955-959.

24. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nat Protoc*. 2010;5(9):1564-1573.

25. Clarke GM, Anderson CA, Pettersson FH, Cardon LR, Morris AP, Zondervan KT. Basic statistical analysis in genetic case-control studies. *Nature protocols*. 2011;6(2):121-133.

26. Grubb A, Blirup-Jensen S, Lindstrom V, et al. First certified reference material for cystatin C in human serum ERM-DA471/IFCC. *Clin Chem Lab Med*. 2010;48(11):1619-1621.

27. Stevens LA, Coresh J, Schmid CH, et al. Estimating GFR using serum cystatin C alone and in combination with serum creatinine: A pooled analysis of 3,418 individuals with CKD. *Am J Kidney Dis*. 2008;51(3):395-406.

28. Levey AS, de Jong PE, Coresh J, et al. The definition, classification, and prognosis of chronic kidney disease: A KDIGO controversies conference report. *Kidney Int*. 2011;80(1):17-28.

29. Inker LA, Levey AS, Tighiouart H, et al. Performance of glomerular filtration rate estimating equations in a community-based sample of blacks and whites: The multiethnic study of atherosclerosis. *Nephrology Dialysis Transplantation*. 2017:gfx042.

30. Böger CA, Gorski M, Li M, et al. Association of eGFR-related loci identified by GWAS with incident CKD and ESRD. *PLoS Genet*. 2011;7(9):e1002292.

31. Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM. Pitfalls of predicting complex traits from SNPs. *Nature Reviews Genetics*. 2013;14(7):507-515.

32. Gorski M, Tin A, Garnaas M, et al. Genome-wide association study of kidney function decline in individuals of european descent. *Kidney Int*. 2015;87(5):1017-1029.

33. Boger CA, Chen MH, Tin A, et al. CUBN is a gene locus for albuminuria. *J Am Soc Nephrol*. 2011;22(3):555-570.

34. Teumer A, Tin A, Sorice R, et al. Genome-wide association studies identify genetic loci associated with albuminuria in diabetes. *Diabetes*. 2016;65(3):803-817.

35. Ellis JW, Chen MH, Foster MC, et al. Validated SNPs for eGFR and their associations with albuminuria. *Hum Mol Genet*. 2012;21(14):3293-3298.

36. Zheng J, Erzurumluoglu AM, Elsworth BL, et al. LD hub: A centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics*. 2017;33(2):272-279.

37. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 2015.

38. Kottgen A. Genome-wide association studies in nephrology research. *Am J Kidney Dis*. 2010;56(4):743-758.

39. Wyss M, Kaddurah-Daouk R. Creatine and creatinine metabolism. *Physiol Rev*. 2000;80(3):1107-1213.

40. Reznichenko A, Snieder H, van den Born J, et al. CUBN as a novel locus for end-stage renal disease: Insights from renal transplantation. *PLoS One*. 2012;7(5):e36512.

Genome-wide association scan of serum urea in European populations identifies two novel loci

CHL Thio, A Reznichenko, PJ van der Most, Z Kamali, A Vaez, Lifelines Cohort Study group, JH Smit, BW Penninx, T Haller, E Mihailov, A Metspalu, J Damman, MH de Borst, P van der Harst, N Verweij, GJ Navis, RT Gansevoort, IM Nolte, H Snieder

Am J Nephrol. 2019;49(3):193-202.

CHAPTER

6

ABSTRACT

Introduction. Serum urea level is a heritable trait commonly used as a diagnostic marker for kidney function. GWAS in East-Asian populations identified a number of genetic loci related to serum urea but there is a paucity of data for European populations.

Methods. We performed a two-stage meta-analysis of GWASs on serum urea in 13,312 participants, with independent replication in 7379 participants of European ancestry.

Results. We identified six genome-wide significant SNPs in or near six loci, of which two were novel (*POU2AF1* and *ADAMTS9-AS2*). Replication of East-Asian and Scottish data provided evidence for an additional eight loci. SNPs tag regions previously associated with anthropometric traits, serum magnesium, and urinary albumin-to-creatinine ratio, as well as expression quantitative trait loci for genes preferentially expressed in kidney and gastro-intestinal tissues.

Conclusions. Our findings provide insights in the genetic underpinnings of urea metabolism, with potential relevance to kidney function.

INTRODUCTION

Serum urea is a diagnostic marker of renal function widely used in clinical practice. Urea is eliminated by the kidneys into urine as waste product of protein metabolism. The net serum urea concentration therefore reflects the excretory capacity of the kidney and elevated values are interpreted as reduced kidney function. Serum urea (or blood urea nitrogen, BUN, when only the nitrogen part is assayed), along with creatinine, is the most frequently requested measurement of kidney function in the assessment of patients with kidney disease. These two markers are not equivalent in estimation of kidney function, and in some conditions (peritoneal dialysis, heart failure) serum urea is considered to be superior to creatinine[1-3]. Alternatively to single-marker use, urea-to-creatinine (or BUN-to-creatinine, respectively) ratio can be used for differential diagnosis of acute kidney injury (prerenal, postrenal, or renal) when one marker is disproportionally elevated or lowered relative to the other[4-6].

Serum urea concentration is highly variable (reference range 1.8-7.1 mmol/L), and besides kidney function, it also depends on hydration status, metabolic rate, dietary protein intake, medication use, liver and cardiac function[5, 6]. Genetic factors may also play a role: one twin study estimated heritability for serum urea concentration to be 44%[7], indicating a contribution of genetic factors to the inter-individual variability of this measure. Furthermore, genome-wide association studies (GWAS) on BUN in East-Asians reported SNP associations at 13 loci[8-11]. For Europeans, there is paucity of data. A recent single-cohort study in the UK did not find any significant associations with urea levels[12], while in a Scottish single-cohort study (N=19,293), five genetic variants were associated with urea[13]. These findings are yet to be replicated in other European cohorts. Concurrently, multiple GWASs in individuals of European descent identified a number of loci associated with serum creatinine and creatinine-based indices of kidney function[14-18]. The genetics underlying urea and creatinine are expected to overlap, because, to a large extent, the serum concentration of both are influenced by kidney function. The studies in East-Asians confirm this notion as they reported *MPPED2-DCDC5* to be associated with both urea and creatinine[10], thus suggesting involvement of this gene with regulation of kidney function. Furthermore, family data from the UK show a positive genetic correlation between urea and creatinine (r_g=0.56)[12]. The existence of exclusively urea-associated loci is also plausible, given that serum levels are not only dependent on kidney function. Identifying these loci will help explain a proportion of kidney function-independent inter-individual variability in urea levels

in the general population and ultimately will provide insight into pathways and regulating mechanisms involved in this metabolic compound.

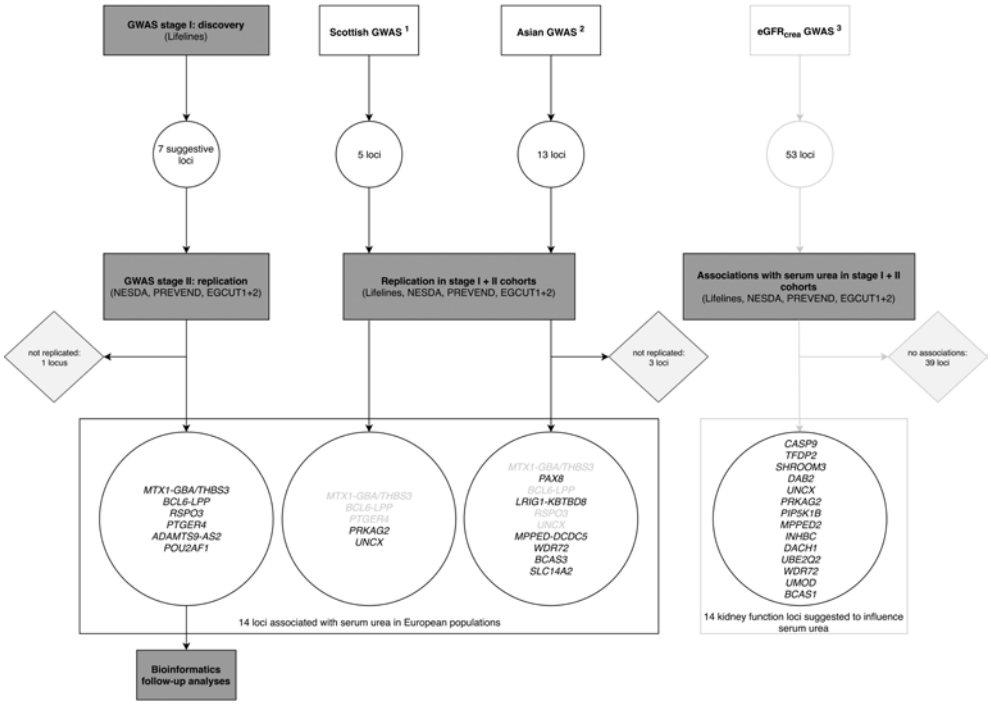
We therefore aimed to identify genetic loci influencing serum urea concentrations in populations of European ancestry. In addition, we compared our results with previous findings from East-Asian and Scottish studies to identify shared loci for serum urea.

METHODS

Study design

An overview of the study design is provided in **Figure 1**. Our strategy consisted of a number of steps. First, we performed a two-stage meta-analysis of GWAS to identify SNPs associated with serum urea. Second, we performed a replication study of loci identified in previous GWAS in East-Asian and Scottish populations. Third, we examined whether known eGFRcrea loci were also associated with serum urea. Furthermore, we conducted bioinformatics follow-up analyses on identified SNPs to identify candidate loci. Each step is detailed below.

Figure 1. Design and results of the present study. Genetic loci in GREY typefont indicate that these loci overlap between GWAS studies on serum urea/BUN.



Study population

Stage I discovery analyses were performed in 13,312 subjects from the Lifelines Cohort Study. Stage II replication testing was performed in 7379 subjects from the PREVEND (N=3387), NESDA (N=2523), EGCUT1 (N=712), and EGCUT2 (N=757) cohorts (**Supplementary Note 1**).

The Lifelines Cohort Study is a multidisciplinary prospective population-based cohort study with a unique three-generation design that examines health and health-related behavior of 165,729 participants living in the north-eastern region of the Netherlands (www.lifelines.nl/researcher). Participants were recruited from November 2006 to December 2013. Eligible individuals were invited through their general practitioner or through participating family members. Additionally, there was the option to self-register. The recruitment and data collection, as well as the representativeness of the data have been described in detail elsewhere[19,20]. Of the 165,729 participants, 15,368 presumably unrelated, oldest members of their respective families, were genotyped (details below). The Lifelines Cohort Study was conducted according to the guidelines in the Declaration of Helsinki and all procedures involving human subjects were approved by the Medical Ethical Committee of the University Medical Center Groningen. Written informed consent was obtained from all participants during their visit at one of the research centers.

Genotyping, quality control, and imputation

A total of 15,368 individuals of the Lifelines Cohort Study were genotyped using the Illumina HumanCytoSNP-12 array and called using GenomeStudio (San Diego, CA, USA). Only autosomal single nucleotide polymorphisms (SNPs) were used in this study. SNPs were excluded when the call rate was <95%, when the minor allele frequency (MAF) was <1%, or when the p-value of the Hardy-Weinberg equilibrium (HWE) test was $<10^{-6}$. Samples were removed when the call rate was <95%, when there was a sex mismatch between database and genotypes, when the heterozygosity deviated >4 SD from the mean heterozygosity over all samples, when it was a first-degree relative to a sample that had a higher call rate, or when non-Caucasian ancestry was likely. After quality control, a total of 268,407 SNPs and 13,385 samples remained. The resulting dataset was phased using MACH[21] and imputed using Minimac[22] with the HapMap Phase 2 CEU haplotypes[23] as reference set. SNPs with an imputation quality $r^2<0.3$ or a

MAF<1% were excluded after imputation. The resulting number of SNPs available for analysis was 1.99 x10^6. The procedure for genotyping, quality control, and imputation of the replication cohorts is described in **Supplementary Note S1**.

Phenotype measurement in Lifelines

At the baseline examination, the participants in the study were asked to fill in a questionnaire before the visit. During the visit, a number of investigations were conducted and blood and 24h-urine samples were taken. A total of 13,385 genotyped participants were included into the present study. The final number of individuals analyzed for serum urea was 13,312 after excluding subjects with extreme values of urea deviating >4 standard deviations (SD) from the mean. Serum urea measurements were performed with an ultraviolet kinetic assay on a Roche Modular. Serum creatinine was measured by an enzymatic method, IDMS traceable on a Roche Modular (Roche, Mannheim, Germany). We estimated eGFRcrea with the 4-variable Modification of Diet in Renal Disease (MDRD) Study equation[24]. Body-mass index (BMI, kg/m^2) was calculated by dividing weight(kg) by squared height (m^2).

Statistical analysis

Three GWASs on serum urea were performed. In the first GWAS, a linear regression for each SNP was performed using an additive SNP model adjusting for age, age^2, sex, body mass index (BMI), and the first ten principal components to adjust for population stratification using PLINK[25]. In the second GWAS, $\log_{10}$-transformed eGFRcrea, was added to the model. In a third GWAS, we adjusted for serum creatinine instead of *log*eGFRcrea. In addition to these three GWAS, we performed sex-stratified analyses. Next, the GWAS results were checked for quality using the QCGWAS package in R[26]. For each GWAS, suggestive SNPs (p-value <10^{-6} in Stage I analyses) were clumped for linkage disequilibrium (LD; r^2>0.1) using pairwise LD checking in SNAP[27] to identify independent index SNPs. These suggestive index SNPs were taken forward to Stage II replication.

The same linear regression analyses as described above were applied to the suggestive SNPs identified in the discovery sample in each of the four replication cohorts separately. The replication results of these SNPs were meta-analyzed using an inverse variance fixed-effects meta-analysis as implemented in the software package GWAMA[28]. A SNP was considered replicated with a one-sided

p-value <0.05 (i.e. same direction of effect), and with significance at the genome-wide level in combined Stage I+II samples ($p<5 \times 10^{-8}$).

Finally, we also sought to replicate 20 SNPs at 13 genetic loci previously identified in GWASs of East-Asian samples[8-11], as well as five SNPs at five loci identified in a Scottish sample[13]. The replication results of these 25 SNPs were meta-analyzed using an inverse variance fixed-effects meta-analysis as implemented in the software package GWAMA[28]. We used all five cohorts (i.e. Lifelines, NESDA, PREVEND, EGCUT1+2) for these analyses. We considered a SNP replicated at a one-sided p<0.05.

Secondary analyses and Bioinformatics

Associations with kidney function

We meta-analyzed associations of 53 known kidney function SNPs[17] with serum urea in all Stage I+II cohorts. Conversely, to examine associations of our six index SNPs with kidney function, we searched publicly available summary data from the same meta-analysis of GWAS on eGFRcrea[17]. At a one-sided p<0.05, we tested whether variants genome-wide significantly associated with lower eGFRcrea were associated with higher urea, and whether SNPs genome-wide significantly associated with higher urea were associated with lower eGFRcrea.

Proportion of phenotypic variance explained

We estimated the proportion of phenotypic variance explained in the NESDA cohort by regressing serum urea level on a weighted genetic risk score (GRS) comprising the effects of all six index SNPs, of the six index SNPs +11 independent SNPs from the Scottish and East-Asian studies, and of the 53 eGFRcrea SNPs. These analyses were performed using PLINK[25] and R[29] on independent SNPs (ldlink.nci.nih.gov) using the effect sizes from the discovery sample (our six index SNPs) or from literature as weights.

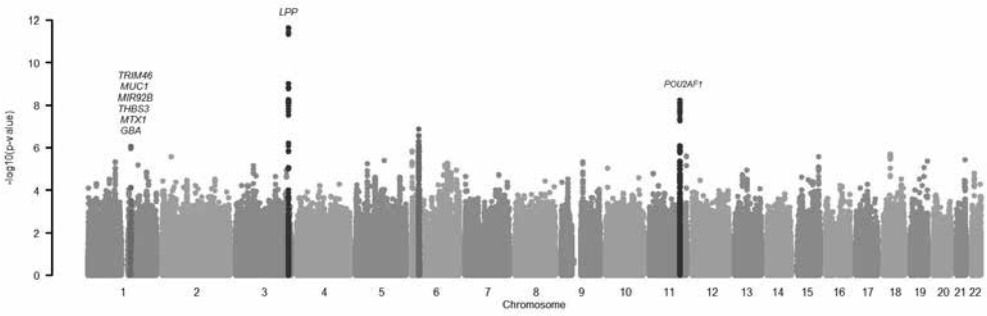
Bioinformatics characterization of the replicated SNPs

We examined functionality (i.e. non-synonymous SNPs and expression quantitative trait loci, eQTL) of the identified index SNPs. To this end, we first converted the positions of all replicated index SNPs to NCBI build 37. We then used the 1000 Genomes Project phase3 release[30] of variant calls to find proxy SNPs in moderate ($r^2>0.5$) and high LD ($r^2>0.8$) with our index SNPs. This dataset is based on the 2013-

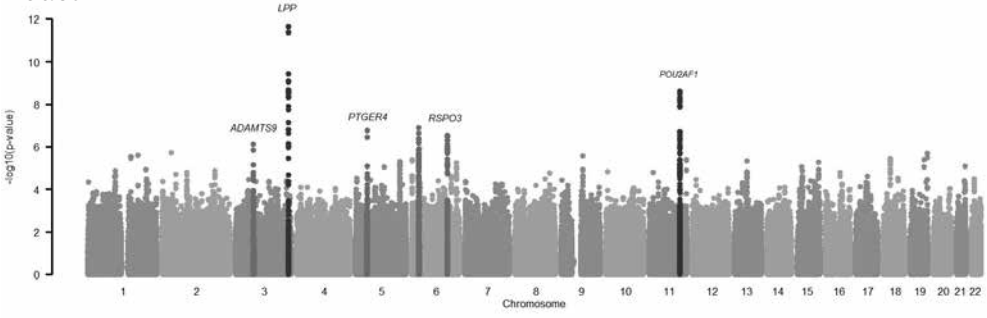
05-02 sequence freeze and alignments. We used version v5a (Feb. 20th, 2015), including the 503 subjects of European ancestry. We used ANNOVAR (version 16 July 2017) (annovar.openbioinformatics.org)[31] for annotation of the index SNPs. We queried PolyPhen-2 (genetics.bwh.harvard.edu/pph2/)[32] to assess whether effects of non-synonymous SNPs were predicted to be malignant. Furthermore, we performed a lookup of the index and proxy SNPs in the GWAS catalog[33] to ascertain whether these SNPs were previously associated with other phenotypes. Genes close to the six index SNPs were followed-up for local expression (*cis*eQTL) in various tissues based on publicly available transcriptomics data: Human Protein Atlas (www.proteinatlas.org)[34], GTEx Portal (www.gtexportal.org/)[35], and blood tissue (genenetwork.nl/bloodeqtlbrowser/)[36]. Furthermore, we examined eQTLs in donor kidney tissue in TransplantLines (detailed description of data and methods in **Supplementary Note 11**)[37, 38].

Figure 2. Manhattan plots of stage I GWAS for serum urea level. The x-axis represents chromosomal position. The y-axis represents two-sided significance on the –log10 scale. Dark grey indicates genome-wide significant hit ($p<5 \times 10^{-8}$), grey indicates suggestive hit ($5 \times 10^{-8} \leq p < 1 \times 10^{-6}$).
Model 1: adjusted for age, age^2, sex, BMI, principal components 1-10
Model 2: Model 1 + logeGFRcrea

Model 1



Model 2



RESULTS

Meta-analysis results

Manhattan plots of stage I for models 1 and 2 are shown in **Figure 2**. Regional association plots, showing location and significance of top hits for models 1 and 2 relative to known loci, are shown in **Supplementary Figure S3**. Risk of bias due to population stratification was assessed and considered acceptable (=1.05) (**Supplementary Figure S4**).

For models 1 and 2, seven index SNPs were at least suggestive ($p<1x10^{-6}$) in stage I. Of these seven SNPs, rs17586946 on chromosome 6 was only suggestive in the combined Stage I+II samples ($p=1.4x10^{-7}$) and hence not replicated. **Table 1** shows results of the remaining six SNPs. For model 1, we replicated three SNPs (rs914615, rs4686914, rs2003313) at three genomic loci, significantly associated with serum urea at the genome-wide level ($p<5x10^{-8}$) in the combined Stage I+II samples. In the second, *log*eGFRcrea-adjusted model, two SNPs from model 1 (rs4686914 and rs2003313) were again identified, while in addition three other SNPs (rs998394, rs11954639, rs2503107) were identified and replicated with genome-wide level significance. One SNP (rs914615) did not reach suggestive significance of $p<1x10^{-6}$ after *log*eGFRcrea adjustment ($p=2.9x10^{-6}$) and therefore deemed non-significant for this model. A third, serum creatinine adjusted model, yielded essentially the same results as the *log*eGFRcrea-adjusted model **(Supplementary Figure 2a, Supplementary Table S5**).

Sex-stratified analysis yielded no additional loci: 1) we found no significant associations in females-only models, and 2) in males-only models, we identified two additional SNPs (rs9860469 and rs9820812) in high linkage disequilibrium (LD) ($r^2=0.70$ and $r^2=1.0$, respectively) with a SNP already identified in models 1-2 (rs4686914) (**Supplementary Figure S2B**). Effects of rs4686914 and rs11954639 were stronger in men (Supplementary Table S6).

Replication of previously reported urea loci

We replicated 10 out of 13 East-Asian loci[8-11] at a one-sided $p<0.05$ (**Supplementary Table S7a**). SNPs at three loci (*MECOM, C12orf51, GNAS*) were not replicated in the present study. All five Scottish loci[13] were replicated (**Supplementary Table S7b**). In total, 14 loci are now confirmed for Europeans (**Figure 3**).

Table 1. Replicated SNP associations with serum urea

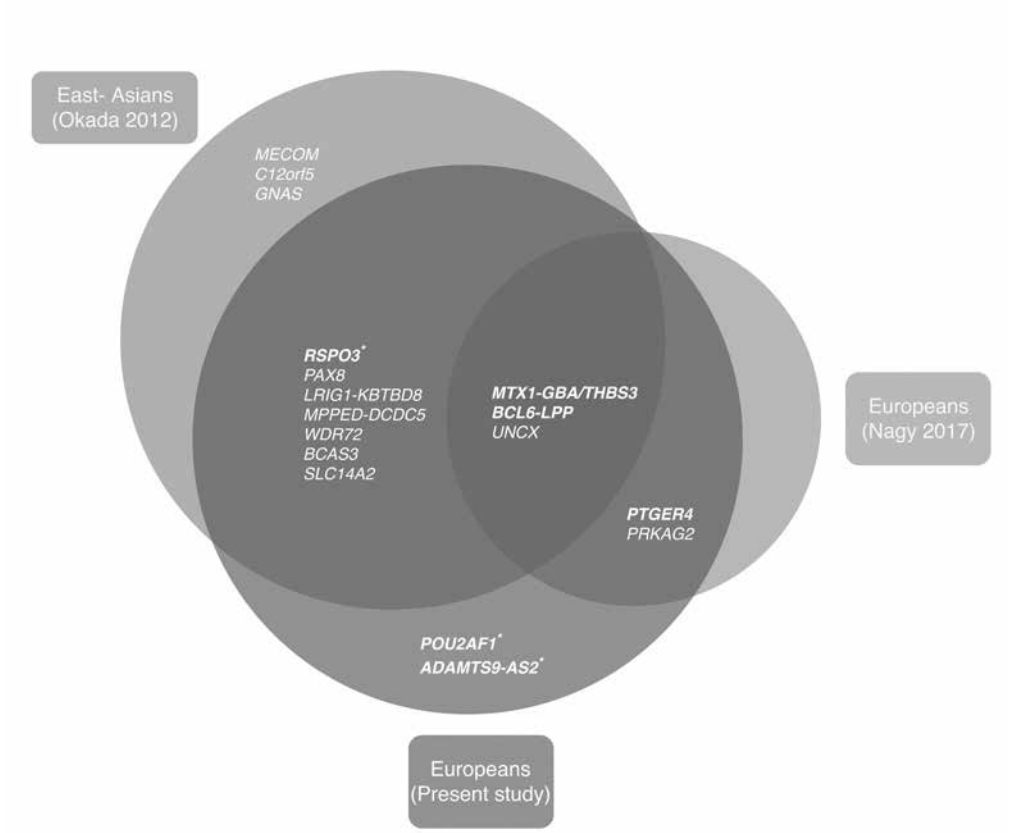
SNP ID	Chr	Position (bp)[a]	Type	Nearest gene	Effect/Non effect allele (EAF)[b]	Model	Stage I (LifeLines)				Stage II (PREVEND, NESDA, EGCUT1+2)				Stage I+II				I² %
							B	SE	p	N	B	SE	p	N	B	SE	p	N	
rs914615	1	153442516	intronic	*THBS3*	A/G (0.476)	1	0.070	0.014	8.9E-07	13312	0.065	0.020	1.3E-03	7379	0.068	0.012	**4.3E-09**	20689	0.0
						2	0.064	0.014	2.9E-06	13311	0.063	0.020	1.2E-03	7335	0.064	0.011	**1.3E-08**	20646	0.0
rs4686914	3	189200234	intergenic	*LPP*	T/C (0.308)	1	-0.110	0.016	2.4E-12	13312	-0.101	0.021	2.2E-06	7378	-0.107	0.013	**2.6E-17**	20690	0.0
						2	-0.106	0.015	2.3E-12	13311	-0.098	0.021	2.1E-06	7334	-0.103	0.012	**2.3E-17**	20645	0.0
rs998394	3	64776227	ncRNA/intronic	*ADAMTS9-AS2*	A/G (0.458)	1	-0.063	0.014	7.3E-06	13312	-0.049	0.020	1.4E-02	7379	-0.058	0.011	3.7E-07	20691	0.0
						2	-0.067	0.014	7.5E-07	13311	-0.058	0.019	2.2E-03	7335	-0.064	0.011	**7.1E-09**	20646	0.0
rs11954639	5	40710736	intergenic	*PTGER4*	T/C (0.071)	1	-0.165	0.037	5.8E-06	13312	-0.170	0.040	2.4E-05	7379	-0.168	0.027	**6.1E-10**	20691	0.0
						2	-0.185	0.035	1.8E-07	13311	-0.182	0.039	2.9E-06	7335	-0.183	0.026	**2.3E-12**	20646	0.0
rs2503107	6	127505069	intronic	*RSPO3*	C/A (0.449)	1	-0.075	0.017	8.6E-06	13312	-0.051	0.020	1.2E-02	7377	-0.065	0.013	4.9E-07	20689	0.0
						2	-0.084	0.016	2.9E-07	13311	-0.056	0.020	4.2E-03	7333	-0.072	0.013	**8.1E-09**	20644	18.0
rs2003313	11	110709203	intergenic	*POU2AF1*	T/A (0.448)	1	-0.088	0.015	6.0E-09	13312	-0.048	0.020	1.7E-02	7377	-0.073	0.012	**1.3E-09**	20691	60.6
						2	-0.087	0.015	2.5E-09	13311	-0.055	0.019	4.3E-03	7333	-0.075	0.012	**9.5E-11**	20644	43.2

Meta-analysis of associations obtained from linear regressions of replicated SNPs with serum urea level assuming additive effects of alleles. Estimates of B and se are presented in mmol/L.
Abbreviations: B, unstandardized regression coefficient; Chr, chromosome; bp, basepair; EAF, effect allele frequency; I^2, heterogeneity statistic; SE, standard error; SNP, single nucleotide polymorphism;

[a] position based on NCBI b36/hg18
[b] EAF in the complete sample (Stage I + II)
˙not suggestive (p≥1E-06) in Stage I for this model

Model 1: adjusted for age,age², sex, body-mass index, principal components 1-10
Model 2: model 1 + log_{10} eGFRcrea

Figure 3. Overview of all 17 currently identified genetic loci. Overlap indicates replication in present study. The six ***BOLD*** loci are genome-wide significant ($p<5 \times 10^{-8}$) in the present study; all other loci in overlapping areas were replicated in the present study at a one-sided p<0.05.
* Novel loci for European populations.



Secondary analyses and Bioinformatics

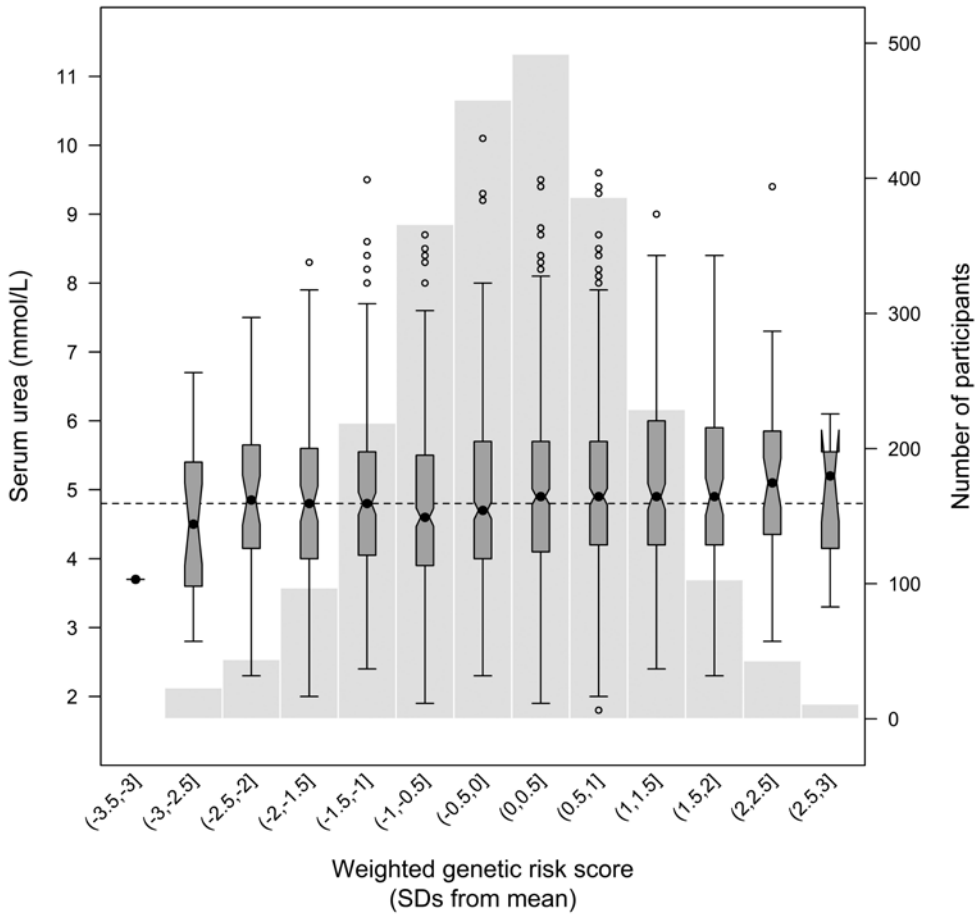
Associations with kidney function

One index SNP (rs2003313) was significantly associated with kidney function, though not in the expected direction (**Supplementary Figure and Table S8a-b**). rs914615 and rs2503107 were borderline significantly associated with kidney function (p=0.095 and p=0.085) in the expected direction. Conversely, 53 known eGFRcrea SNPs[17] were examined for potential associations with serum urea levels in all Stage I+II cohorts. After meta-analysis, 14 of the 53 SNPs were significantly associated with serum urea levels (**Supplementary Figure and Tables S9a-c**), more than could be expected through random chance alone (binomial distribution, 14/53, =0.05, $p=1.98 \times 10^{-7}$).

Proportion of phenotypic variance explained in the NESDA cohort

A GRS comprising all six index SNPs explained a small, but significant proportion of 0.43-0.45% of phenotypic variation in NESDA (**Supplementary Table S10**). This increased to 0.45-0.56% when 11 independent SNPs were added from the Scottish and East-Asian studies. A weighted GRS comprised of all 17 SNPs showed a modest but significant linear trend ($p<2.3x10^{-4}$) in urea levels (**Figure 4**).

Figure 4. Distribution of serum urea levels. Boxplots of serum urea levels (mmol/L) by categories of a weighted genetic risk score comprised of all 17 currently identified serum urea SNPs in the NESDA cohort (N=2472). The black dots represent the medians; the grey boxes represent the observations between the 25th and the 75th percentile; the whiskers represent (at maximum) 1.5 times the interquartile range; the notches represent the 95%CI of the median. In the rightmost boxplot, the notches extend to outside the box due to its wide 95%CI. The underlying light gray histogram represents the population distribution of the genetic risk score; its bell shape approximates a normal distribution. The dashed horizontal line depicts the median serum urea level in the NESDA cohort (4.8 mmol/L).



However, we observed no clinically relevant differences in serum urea between extremes of this GRS. The 53 SNPs identified to be associated with serum creatinine by the CKDGen consortium explained 0.18% of the variance in serum urea (p=0.02), but significance of this effect disappeared when correcting for *log*eGFRcrea or serum creatinine.

Bioinformatics characterization of the index SNPs

Our analyses returned 345 SNPs in at least moderate LD (r^2>0.50), of which 173 in at least high LD (r^2>0.80) and 49 in perfect LD (r^2=1). rs914615 is linked with two non-synonymous SNPs: rs760077 (*MTX1*), and rs4745 (*EFNA1*), both of which are predicted to be benign[32]. A number of proxy SNPs in high LD (r^2>0.8) with the index SNPs were reported in the literature as associated with other kidney-function or metabolically-relevant traits such as serum magnesium level and anthropomorphic traits. rs914615 was previously found associated with urinary albumin-to-creatinine ratio in diabetic subjects[39] (**Supplementary Table S13**). Using eQTL data publicly available from GTEx Portal, we found associations of three SNPs with gene expression in various tissues, and predominantly in gastro-intestinal tissues (Supplementary Table S14): rs914615 with expression of numerous genes, among others *EFNA1, MTX1, MUC1,* and *THBS3*; rs2003313 with *COLCA1* and *COLCA2*; and rs11954639 with *RPL37*. In whole blood, SNP rs914615 was associated with expression of *THBS3*, *ADAM15*, *KRTCAP2* (**Supplementary Table S15**). In kidney biopsy specimens, we found an association of the A allele of rs914615 with decreased mucin gene (*MUC1*) expression (**Supplementary Table S16**).

DISCUSSION

In this meta-analysis of GWAS in European populations, we identified six index SNPs at six genomic loci (in *THBS3, ADAMTS9-AS2, RSPO3,* or near *LPP, PTGER4, and POU2AF1*) that were associated with serum urea levels at a genome-wide significant level. Of these six index SNPs, two (near *POU2AF1* and in *ADAMTS9-AS2*) are completely novel associations with urea, i.e. not previously identified in either the East-Asian or Scottish studies. Three SNPs tag regions (*THBS3, LPP, and RSPO3*) previously identified in East-Asians. SNP rs11954639 near *PTGER4* is in high LD with a SNP previously identified in Scottish GWAS. Follow-up analysis of the six index SNPs yielded potential roles of a number of loci in urea metabolism.

In addition to our main meta-analysis, we examined 20 SNPs at 13 genetic loci previously associated with BUN in East-Asians[8-11]. Of these 20 SNPs, we replicated 15 at a one-sided p<0.05, confirming 10 previously identified loci (*MTX1-GBA*, *PAX8*, *BCL6-LPP*, *LRIG1-KBTBD8*, *RSPO3*, *UNCX*, *MPPED-DCDC5*, *WDR72*, *BCAS3*, and *SLC14A2*) but not *MECOM*, *C12orf51*, and *GNAS*. Of note, we replicated SNPs at the *SLC14A2* locus, a gene that encodes a renal tubular urea transporter (RefSeq release 89)[40]. Furthermore, we confirmed SNP associations at *MTX1, RP11-115 J16.1*, *PRKAG2, UNCX*, and an intergenic region near *PTGER4*, that were identified in a single-cohort GWAS in 19,293 Generation Scotland participants[13]. After replication, SNPs at 14 loci now have confirmed associations with serum urea in Europeans. SNPs tagging *PTGER4, PRKAG2, ADAMTS9-AS2,* and *POU2AF1* were specific to European studies, likely due to considerably lower minor allele frequencies in East-Asians (0%, 0%, 16%, and 12%, respectively) compared with Europeans (7%, 30%, 46%, 44%) according to the 1000G phase 3 East-Asian (EAS) and European (EUR) reference sets[30].

GWAS of biomarkers that are excreted through the kidney may be confounded by kidney function[41]. We therefore examined the effect of kidney function on SNP associations by running both unadjusted models and *log*eGFRcrea-adjusted models. Associations of two SNPs (rs4686914, rs2003313) were unaffected by this adjustment, thus are suggested to affect urea levels not through kidney function but through other mechanisms. Associations of three SNPs (rs998394, rs11954639, rs2503107) were only significant in the *log*eGFRcrea-adjusted model, indicating positive confounding/suppression, i.e. genetic effects were masked by kidney function. Associations of one SNP (rs914615) diminished after *log*eGFRcrea adjustment, suggesting that the effect of this SNP on serum urea is (partly) confounded or mediated through kidney function. In the following paragraphs, we discuss the two novel loci.

We report a novel association of urea with rs2003313, a SNP on chromosome 11 in an intergenic region near *POU2AF1*. We queried the GWAS catalog to find other phenotypes associated with this SNP, and SNPs in LD, r^2>0.50); however, we found none. eQTL analysis in GTEx[35] yielded significant associations of rs2003313 with expression of *COLCA2* and *COLCA1* (aliases *C11orf93* and *C11orf92*, respectively) in colon, esophagus, spleen, tibial artery and nerve, and adipose tissue. Protein function of *COLCA2* is currently unknown. *COLCA1* encodes a transmembrane protein of granular structures, such as crystalloid eosinophilic granules and other

granular organelles[40], with preferential expression in stomach, urinary bladder, and prostate. Both *COLCA2* and *COLCA1* have previously been associated to colorectal cancer[42]. Relevance of this locus to serum urea is unclear, and may be explored in future study. Against expectations, the T allele of rs2003313 was associated with lower serum urea in the present study, and with lower eGFRcrea in CKDGen data[17]. Whether this is due to unmeasured confounding or some unknown biological factor may be explored in future study. Potential biological mechanisms may be explored in future study. Of note, moderate heterogeneity was observed (I^2: 43-61%) with diminution of effect size in the replication phase, possibly indicative of Winner's curse[43], i.e. the effect of this SNP may be overestimated. Nonetheless, the strong significance of the combined meta-analysis of this locus indicates it is a non-spurious signal.

A second novel SNP is rs998394 on chromosome 3. Although in relative proximity (distance ~2Mb) to SNPs (near *LRIG1-KBTBD8*) previously identified in East-Asian GWAS on BUN, these are not in linkage disequilibrium (r^2=0.0); we thus consider this SNP independent and therefore a novel finding. rs998394 is located in *ADAMTS9-AS2*, a long non-coding RNA that is an antisense transcript of *ADAMTS9*. The protein encoded by *ADAMTS9* is a member of the ADAMTS (a disintegrin and metalloproteinase with thrombospondin motifs) protein family. Members of this family have been implicated in the cleavage of proteoglycans, the control of organ shape during development, and the inhibition of proteoglycans[40]. *ADAMTS9* is localized to chromosome region 3p14.3-p14.2, an area known to be lost in hereditary renal tumors[44]. *ADAMTS9* has previously been associated with anthropomorphic traits[45, 46] and type 2 diabetes mellitus[47].

Loci tagged by the other four index SNPs are discussed in Supplementary Note S12. Briefly, we found potential roles of *MUC1* and *PTGER4* in urea metabolism and/or kidney function.

Sex-stratified analysis yielded no additional loci, although a marked difference in effect size was observed between men and women for rs4686914 and rs11954639. This is suggestive of gender-specific mechanisms of urea metabolism which may be investigated in future study.

Fourteen out of 53 (26%) known eGFRcrea loci were associated (one-sided p<0.05) with serum urea levels in our discovery cohort, more than could be expected through random chance alone. Furthermore, a GRS based on these loci was modestly but significantly associated with serum urea, supporting the notion of genetic overlap between the two traits. Previously, Okada et al observed associations of *MPPED-DCDC5, BCAS3, WDR72,* and *UNCX*, with both creatinine and BUN at the genome-wide level in East-Asians[10], indicating possible pleiotropy. In addition, the present study suggests pleiotropy for *PRKAG2, UNCX,* and *WDR72*, given that these known eGFRcrea loci also associated with serum urea in the present study.

To the best of our knowledge, the present study is the first meta-analysis of GWAS of serum urea in European populations. We were able to report new associations for European populations and confirm known associations from East-Asian studies. However, a genetic risk score combining all currently identified SNPs was only modestly associated with serum urea. Future study may involve imputation to the Haplotype Reference Consortium reference set[48], which due to its higher resolution may yield more precise results. Given the estimated explained variance of the identified SNPs (0.56%), and the estimated heritability of serum urea levels (44%), many of the genetic factors influencing serum urea are still to be found; larger samples are needed to detect these factors. Consequently, the immediate clinical relevance of our findings is limited.

In conclusion, we report the first meta-analysis of GWAS of serum urea levels in European populations. We identified six genomic loci reproducibly associated with serum urea. We are the first to report two SNP associations with urea near *POU2AF1* and in *ADAMTS9-AS2*. The identified regions have possible relevance to urea metabolism, as well as kidney function.

All Supplementary material can be accessed via the following link:
www.karger.com/Article/FullText/496930

REFERENCES

1. Gotch FA: Urea is the best molecule to target adequacy of peritoneal dialysis. *Perit Dial Int.* 2000;20 Suppl 2:S58-64.

2. Aronson D, Mittleman MA, Burger AJ: Elevated blood urea nitrogen level as a predictor of mortality in patients admitted for decompensated heart failure. *Am J Med.* 2004;116(7):466-473.

3. Gotsman I, Zwas D, Planer D, Admon D, Lotan C, Keren A: The significance of serum urea and renal function in patients with heart failure. *Medicine.* 2010;89(4):197-203.

4. Baum N, Dichoso CC, Carlton CE: Blood urea nitrogen and serum creatinine: physiology and interpretations. *Urology.* 1975;5(5):583-588.

5. Hosten AO: BUN and Creatinine; in 3rd, Walker HK, Hall WD, Hurst JW (eds): *Clinical Methods: The History, Physical, and Laboratory Examinations*. Boston, Butterworth Publishers, a division of Reed Publishing, 1990.

6. Dirkx TC, Woodell T: Kidney Disease; in Papadakis MA, McPhee SJ, Rabow MW (eds): *Current Medical Diagnosis & Treatment 2019*. New York, NY, McGraw-Hill Education, 2019.

7. Kettunen J, Tukiainen T, Sarin A, Ortega-Alonso A, Tikkanen E, Lyytikäinen L, et al: Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nat Genet.* 2012;44(3):269.

8. Kamatani Y, Matsuda K, Okada Y, Kubo M, Hosono N, Daigo Y, et al: Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat Genet.* 2010;42(3):210.

9. 9. Kim YJ, Go MJ, Hu C, Hong CB, Kim YK, Lee JY, et al: Large-scale genome-wide association studies in East Asians identify new genetic loci influencing metabolic traits. *Nat Genet.* 2011;43(10):990.

10. Okada Y, Sim X, Go MJ, Wu JY, Gu D, Takeuchi F, et al: Meta-analysis identifies multiple loci associated with kidney function-related traits in East Asian populations. *Nat Genet.* 2012;44(8):904-909

11. Lee J, Lee Y, Park B, Won S, Han JS, Heo NJ: Genome-wide association analysis identifies multiple loci associated with kidney disease-related traits in Korean populations. *PLoS One.* 2018;13(3):e0194044.

12. Prins BP, Kuchenbaecker KB, Bao Y, Smart M, Zabaneh D, Fatemifar G, et al: Genome-wide analysis of health-related biomarkers in the UK Household Longitudinal Study reveals novel associations. *Scientific Reports.* 2017;7(1):11008.

13. Nagy R, Boutin TS, Marten J, Huffman JE, Kerr SM, Campbell A, et al: Exploration of haplotype research consortium imputation for genome-wide association studies in 20,032 Generation Scotland participants. *Genome Med.* 2017;9(1):23.

14. Chambers JC, Zhang W, Lord GM, Van Der Harst P, Lawlor DA, Sehmi JS, et al: Genetic loci influencing kidney function and chronic kidney disease. *Nat Genet.* 2010;42(5):373-375.

15. Kottgen A, Glazer NL, Dehghan A, Hwang S, Katz R, Li M, et al: Multiple loci associated with indices of renal function and chronic kidney disease. *Nat Genet.* 2009;41(6):712-717.

16. Kottgen A, Pattaro C, Boger CA, Fuchsberger C, Olden M, Glazer NL, et al: New loci associated with kidney function and chronic kidney disease. *Nat Genet.* 2010;42(5):376-384.

17. Pattaro C, Teumer A, Gorski M, Chu AY, Li M, Mijatovic V, et al: Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat Commun.* 2016;7.

18. Gorski M, van dM, Teumer A, Chu AY, Li M, Mijatovic V, et al: 1000 Genomes-based meta-analysis identifies 10 novel loci for kidney function. *Scientific Reports.* 2017;7:45040.

19. Scholtens S, Smidt N, Swertz MA, Bakker SJ, Dotinga A, Vonk JM, et al: Cohort Profile: LifeLines, a three-generation cohort study and biobank. *Int J Epidemiol.* 2015 Aug;44(4):1172-80.

20. Klijs B, Scholtens S, Mandemakers JJ, Snieder H, Stolk RP, Smidt N: Representativeness of the LifeLines cohort study *PLoS One.* 2015 Sep 2;10(9)

21. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR: MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol.* 2010;34(8):816-834.

22. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR: Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet.* 2012;44(8):955-959.

23. International HapMap Consortium: A second generation human haplotype map of over 3.1 million SNPs. *Nature.* 2007;449(7164):851.

24. Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D: A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. *Ann Intern Med.* 1999;130(6):461-470.

25. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007 Sep;81(3):559-75.

26. van der Most, Peter J, Vaez A, Prins BP, Munoz ML, Snieder H, Alizadeh BZ, et al: QCGWAS: A flexible R package for automated quality control of genome-wide association results. *Bioinformatics.* 2014;30(8):1185-1186.

27. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'donnell CJ, De Bakker PI: SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics.* 2008;24(24):2938-2939.

28. Mägi R, Morris AP: GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics.* 2010;11(1):288.

29. R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

30. 1000 Genomes Project Consortium: A map of human genome variation from population-scale sequencing. *Nature.* 2010;467(7319):1061-1073.

31. Wang K, Li M, Hakonarson H: ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38(16):e164-e164.

32. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al: A method and server for predicting damaging missense mutations. *Nat Methods.* 2010 Apr;7(4):248-9.

33. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al: The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* 2016;45(D1):D896-D901.

34. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al: Proteomics. Tissue-based map of the human proteome. *Science.* 2015;347(6220):1260419.

35. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al: The genotype-tissue expression (GTEx) project. *Nat Genet.* 2013;45(6):580.

36. Westra H, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al: Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet.* 2013;45(10):1238.

37. Damman J, Bloks VW, Daha MR, van der Most PJ, Sanjabi B, van der Vlies P, et al: Hypoxia and Complement-and-Coagulation Pathways in the Deceased Organ Donor as the Major Target for Intervention to Improve Renal Allograft Outcome. *Transplantation.* 2015;99(6):1293-1300.

38. Wain LV, Vaez A, Jansen R, Joehanes R, van der Most PJ, Erzurumluoglu AM, et al: Novel Blood Pressure Locus and Gene Discovery Using Genome-Wide Association Study and Expression Data Sets From Blood and the Kidney. *Hypertension.* 2017;70:e4-e19.

39. Teumer A, Tin A, Sorice R, Gorski M, Yeo NC, Chu AY, et al: Genome-wide association studies identify genetic loci associated with albuminuria in diabetes. *Diabetes.* 2016;65(3):803-81.

40. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al: Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 2015;44(D1):D733-D745.

41. Wuttke M, Köttgen A: Insights into kidney diseases from genome-wide association studies. *Nat Rev Nephrol.* 2016;12(9):549.

42. Tenesa A, Farrington SM, Prendergast JG, Porteous ME, Walker M, Haq N, et al: Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat Genet.* 2008;40(5):631.

43. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN: Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet.* 2003;33(2):177.

44. Clark ME, Kelner GS, Turbeville LA, Boyer A, Arden KC, Maki RA: ADAMTS9, a novel member of the ADAM-TS/metallospondin gene family. *Genomics.* 2000;67(3):343-350.

45. Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, Steinthorsdottir V, et al: Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat Genet.* 2010;42(11):949.

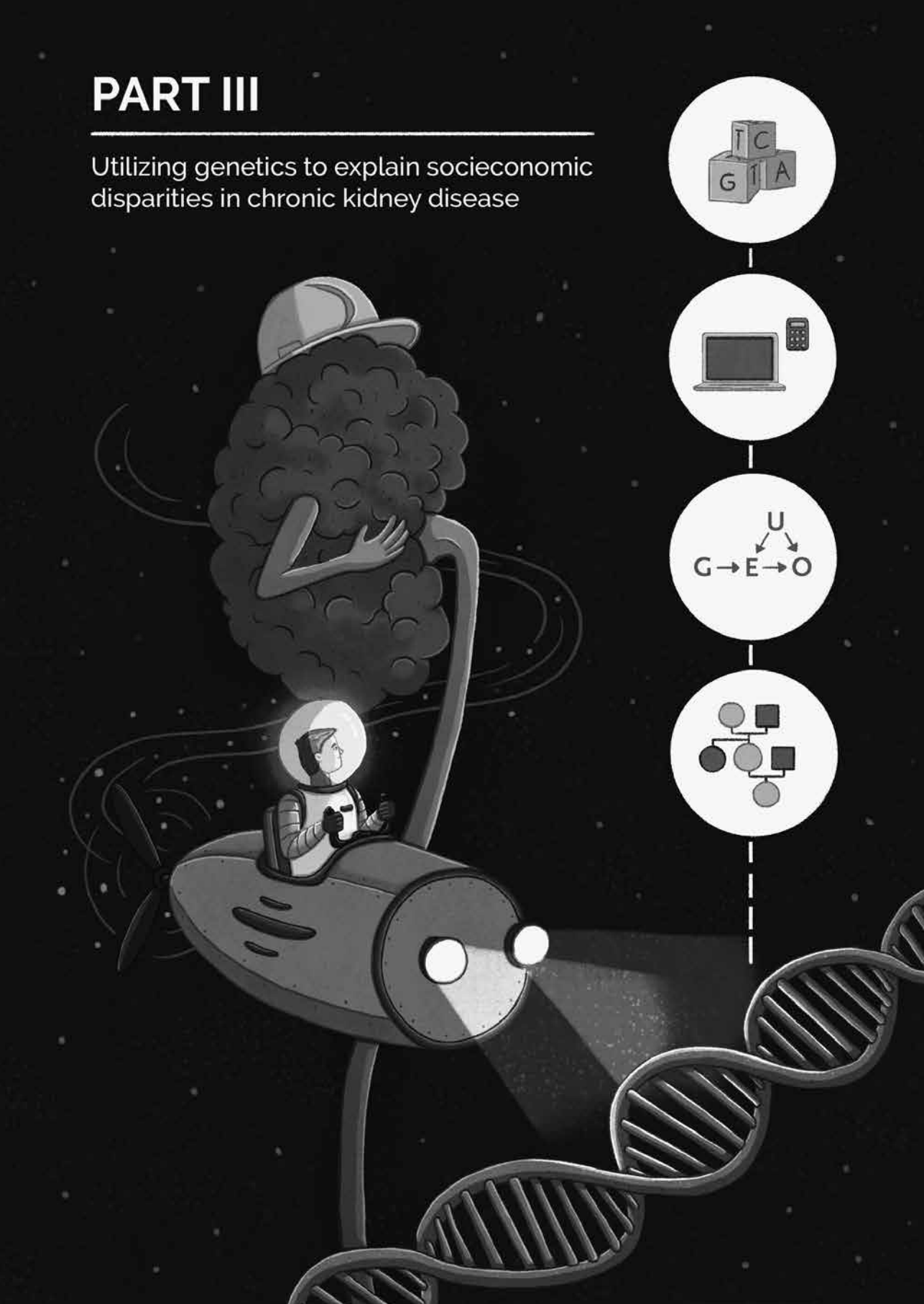
46. Randall JC, Winkler TW, Kutalik Z, Berndt SI, Jackson AU, Monda KL, et al: Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet.* 2013;9(6):e1003500.

47. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, et al: Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet.* 2008;40(5):638.

48. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al: A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet.* 2016;48(10):1279.

PART III

Utilizing genetics to explain socieconomic disparities in chronic kidney disease



The effects of genetic factors, educational attainment, and their interaction on kidney function outcomes

CHL Thio, SKR van Zon, PJ van der Most, H Snieder, U Bültmann, RT Gansevoort

Am J Epidemiol. 2020. In press

CHAPTER

7

ABSTRACT

Introduction. Both genetic predisposition and low educational attainment (EA) are associated with higher risk of chronic kidney disease (CKD). We aimed to examine the joint effects of EA and genetic predisposition, and their interaction on kidney function outcomes.

Methods. We used data from the longitudinal community-based PREVEND Study. Glomerular filtration rate was estimated (eGFR) from serum creatinine and cystatin C using the CKD-EPI equation. For each individual, a linear eGFR trajectory was estimated using linear mixed models. Genotype data on 63 single nucleotide polymorphisms (SNPs), with known associations to eGFR, were used to calculate an allele-weighted genetic score (WGS). Educational attainment was categorized into high, medium, and low EA. Ordinary least squares regression was performed to assess main and interaction effects in cross-sectional and longitudinal analysis, adjusting for age, sex, and renal risk factors (body-mass index, blood pressure, glucose, cholesterol, and smoking).

Results. We included 3597 participants with ~11 years of follow-up. At baseline, a higher WGS and lower EA were independently associated with reduced eGFR and showed additive effects, as an interaction term between the WGS and EA was not significant. In longitudinal analysis, the interaction term was significant ($p=0.036$), and its direction suggested an amplifying effect of low EA on the WGS: those with high genetic risk and low EA had a disproportionately faster rate of eGFR decline relative to those with higher EA. Inclusion of renal risk factors in our models did not change our results.

Conclusion. This is the first study to present evidence of gene-environment interaction between EA and a WGS on eGFR decline, that is not explained by traditional risk factors. These results provide population level insights into the mechanisms underlying socioeconomic disparities in CKD.

INTRODUCTION

Chronic kidney disease (CKD) is a heterogeneous group of disorders characterized by sustained kidney dysfunction and/or signs of kidney damage[1]. CKD is associated with cardiovascular morbidity and all-cause mortality[2]. It may eventually also progress to end-stage kidney disease, necessitating the start of renal replacement therapy. The incidence of CKD is increasing, which poses a major global health challenge[3-5].

Over the last two decades, it has become clear that there is a socioeconomic gradient in CKD risk: low educational attainment (EA), as an indicator of low socioeconomic status (SES), is associated with reduced kidney function (estimated glomerular filtration rate, eGFR) and with higher rates of kidney damage (urinary albumin excretion, UAE)[6,7]. Recent data suggest that indicators of SES including EA are linked with CKD through poor health behaviors (e.g. smoking, diet, sedentary time), higher prevalence of known clinical risk factors (hypertension, diabetes, hypercholesterolemia, obesity), and poor health care access[8,9], each contributing to an environment that is deleterious for kidney health.

In addition to environmental factors, there is strong evidence for a genetic influence on CKD. Familial clustering is observed in CKD[10-13], and heritability of CKD defining traits has been estimated to be 36-75%. Further evidence is provided by genome-wide association studies (GWAS) that identified >60 single nucleotide polymorphisms (SNPs) associated with creatinine-based eGFR (eGFRcrea)[14]. Genetic scores constructed from these SNPs represent a genetic component to kidney function, and thus can be interpreted as a proxy of genetic liability to CKD[15-17].

Some evidence exists, albeit conflicting, that higher education counteracts the genetic risk of diabetes[18,19] and obesity[18,20,21], both important determinants of CKD. Therefore, it is possible that higher education also counteracts genetic risk of CKD, or conversely, that low education amplifies the genetic risk of CKD. Uncovering modifying effects of education on genetic risk may facilitate improved risk stratification based on education and genetics. Furthermore, knowledge of modifying effects of education provides support for public health policies, e.g. in managing downstream effects of low education to improve kidney outcomes. The joint effects of education and genetic factors have not previously been examined

in the context of kidney disease. Thus, our aim was to investigate the interaction between education and genetic predisposition for CKD in the general population. Specifically, we aimed to test the hypothesis that lower EA amplifies genetic risk of reduced kidney function.

METHODS

Study sample and design

We used data from the Prevention of REnal and Vascular ENd stage Disease (PREVEND) Cohort study. PREVEND was initiated to investigate the natural course of increased urinary albumin levels and its association with renal and vascular outcomes. Details of this study have been described elsewhere[22]. Briefly, 8592 individuals, sampled from the general population of Groningen, the Netherlands, underwent an extensive baseline examination between 1997-1998. Four follow-up examinations were completed in 2003, 2006, 2008, and 2012. All subjects gave written informed consent. PREVEND was approved by the medical ethics committee of the University Medical Center Groningen and conducted in accordance with the Helsinki Declaration guidelines. For this study, we included a subset of participants that was genotyped (n=3649). Participants aged <30 years were excluded N=52).

Measurements

Kidney function

Measurement of serum creatinine was performed by an enzymatic method on a Roche Modular analyzer using reagents and calibrators from Roche (Roche Diagnostics, Mannheim, Germany), traceable to isotope dilution mass spectrometry, with intra- and interassay coefficients of variation of 0.9% and 2.9%, respectively. Serum cystatin C concentration was measured by a Gentian cystatin C Immunoassay (Gentian AS Moss, Norway) on a Modular analyzer (Roche Diagnostics). Cystatin C was calibrated directly using the standard supplied by the manufacturer (traceable to the International Federation of Clinical Chemistry Working Group for Standardization of Serum Cystatin C)[23]. The intra- and interassay coefficients of variation were <4.1% and <3.3%, respectively. Serum creatinine and serum cystatin C were determined in a single run to avoid laboratory day-to-day variation. We calculated eGFR from both serum creatinine and serum cystatin C, using the corresponding Chronic Kidney Disease – Epidemiology collaboration (CKD-EPI) equation[24].

Genotyping and genetic risk score calculation

Genotyping details for PREVEND were described previously[17]. Briefly, genotyping was performed on the Illumina CytoSNP12 v2 chip. Variants were imputed to 1000G Phase 1 version 3, using Minimac software. Genetic effects may be confounded by population stratification. Therefore, principal component analysis was performed to reduce dimensionality of the genetic data[25]; the resulting principal components (PCs) represent possible population substructures in PREVEND. In order to remove ethnic outliers, samples with z-score>3 for any of the first five principal components with the highest eigen values were excluded. Samples with call rate <95%, duplicates, and sex discrepancies were also excluded. Markers with call rate >95%, Hardy-Weinberg equilibrium $p \geq 10^{-5}$ and minor allele frequency ≥1% were included. From the resulting GWAS data, we extracted genotypes of 63 SNPs identified in a meta-analysis of GWAS on eGFRcrea in European populations. We constructed a weighted genetic score (WGS) comprising effects of 63 known eGFR SNPs[14]. Per individual, effect alleles were weighted for their published effect sizes and summed. We then standardized the scores by subtracting the population mean score and dividing the score by the population standard deviation. Effect alleles were those reported to associate with lower eGFR, thus a higher WGS reflects genetic predisposition towards lower eGFR.

Educational attainment

Educational attainment (EA) was assessed with self-report questionnaires. EA levels specific to the Netherlands were mapped to the International Standard Classification of Education (ISCED)[26]. We then categorized EA into low (no, primary, basic vocational, and secondary education, corresponding to ISCED levels 0-2), medium (senior secondary vocational and general senior secondary education, ISCED levels 3-4), and high (higher professional and higher academic education, ISCED levels 5-6). ISCED levels were imputed to US years of schooling. High EA was the reference category in all analyses.

Covariates

We adjusted for age, age^2, and sex. To minimize potential confounding by population stratification, we additionally adjusted for the first ten genetic PCs. In longitudinal analyses, we additionally adjusted for baseline eGFR. Furthermore, we explored models that include the renal risk factors, body-mass-index (BMI, weight/height^2), systolic blood pressure (SBP), glucose, total cholesterol, and

smoking status (never smoker, former smoker, current smoker), each measured at baseline. Outliers exceeding four standard deviations (sds) from the mean were excluded.

Statistical analyses

All analyses were performed using R software version 3.5.1[27]. To assess the explained variance of eGFR by the WGS, conditional on age, age^2, sex, and the first ten PCs, $\Delta R^2_{adjusted}$ was computed from nested ordinary least squares regression models using the *lm( )* function from the *stats* R package. We tested associations between the WGS and EA using one-way ANOVA implemented in the *aov( )* function from the *stats* R package.

Cross-sectional analyses, with baseline eGFR as outcome, were performed using ordinary least squares linear regression analysis implemented in the *stats* R package. For longitudinal analyses, we performed a two-step procedure. First, we modelled linear trajectories of eGFR using linear mixed models (LMM) implemented in the *lme4* R package[28], with a random intercept and a random slope for time. Individual trajectories of eGFR change were then extracted and used as outcome variable (i.e., annual eGFR change) in ordinary least squares linear (OLS) regression analysis. For both cross-sectional analyses and longitudinal analyses, six models were constructed with the main effects of the WGS and EA, in addition their interaction term, and varying degrees of covariate adjustment (see **Table 2** for model details). Contribution of the WGS x EA interaction term was assessed using model coefficients for separate EA levels (low EA, medium EA, and the interaction of each with the WGS, with high EA as reference category), and computing the difference in adjusted explained variance ($\Delta R^2_{adjusted}$) between two nested models (with and without interaction term). To assess significance of the overall interaction term, we used an *F*-test using the *anova( )* function from the *stats* R-package, through which we compared model fit between two nested models. We used linear regression models, hence interaction was assessed on the additive scale, and a significant p-value for the interaction term indicates departure from additivity. For all models, we performed a complete-case analysis. We applied a two-sided significance threshold of =0.05 unless otherwise specified.

RESULTS

Baseline characteristics

Baseline characteristics of participants, by categories of EA, are presented in **Table 1**. Lower EA was generally associated with a less favorable renal risk profile (lower eGFR, higher BMI, higher SBP, higher glucose, higher cholesterol, and higher prevalence of smoking).

Table 1. Baseline characteristics overall and by educational attainment.

		Educational attainment		
	Total	**Low**	**Medium**	**High**
N	3597	1673	889	1035
Age (years)	50 [40-60]	55 [46-65]	46 [37-56]	44 [37-51]
Males	52%	49%	56%	53%
eGFR (mL/min/1.73m²)	94.7 ± 17.0	90.5 ± 17.3	97.1 ± 17.0	99.3 ± 14.8
US years of schooling	12.9 ± 5.0	8.5 ± 1.5	13 ± 0	20 ± 0
WGS	0 ± 1	0.02 ± 1.0	-0.02 ± 1.0	-0.01 ± 1.0
Number of effect alleles	62.3 ± 4.9	62.3 ± 4.9	62.3 ± 5.1	62.3 ± 4.8
SBP (mmHg)	129 ± 19.7	133 ± 20	128 ± 20	124 ± 18
Glucose (mmol/L)	4.8 ± 0.8	5.0 ± 0.8	4.7 ± 0.7	4.6 ± 0.6
BMI (kg/m²)	26 ± 4.1	27 ± 4.2	26 ± 4.0	25 ± 3.5
Total cholesterol (mmol/L)	5.7 ± 1.1	5.9 ± 1.1	5.6 ± 1.1	5.4 ± 1.0
Never smoker	27%	23%	26%	36%
Former smoker	37%	37%	38%	37%
Current smoker	35%	40%	36%	27%
Follow-up time (years)	11.0 [4.6 – 11.9]	9.9 [4.2-11.6]	11.1 [4.8-12.2]	11.2 [6.2-12.4]

Data are presented as mean ± standard deviation, median [interquartile range] or percentages. Abbreviations are: eGFR, estimated glomerular filtration rate; WGS, weighted genetic risk score; SBP, systolic blood pressure; BMI, body-mass-index.

We regressed baseline eGFR on the WGS to obtain a crude association. The effect of the WGS on baseline eGFR, was modest but highly significant (B ± se = -1.68 ± 0.29, $R^2_{adjusted}$ = 0.010, p=8.6 x10^{-9}).

No difference in the WGS or risk allele number between categories of EA was observed. We examined the association between EA and the WGS. In **Supplementary Figure 1**, we plot the WGS by categories of EA. The WGS was normally and equally distributed in each EA category. As expected, the mean WGS did not significantly differ between EA categories ($F_{(2, 3594)}$=0.455, p=0.635).

Interaction analyses

Cross-sectional analysis

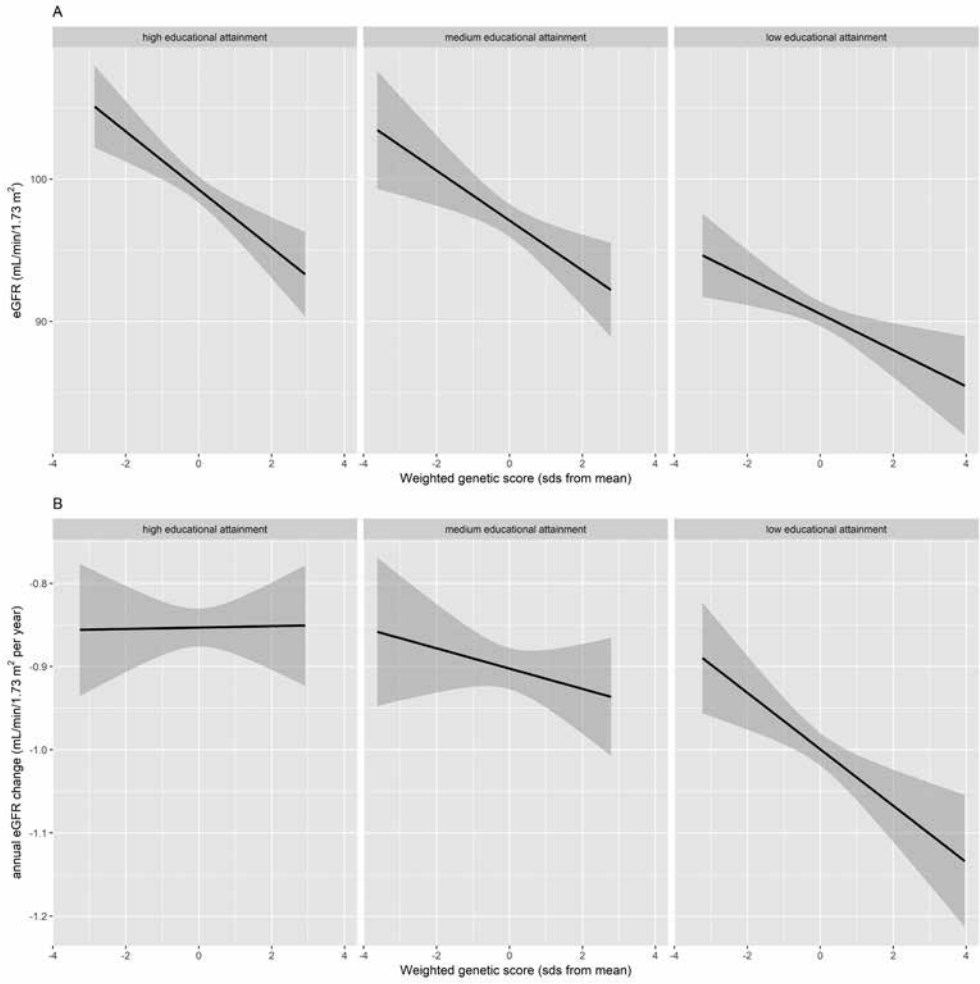
A plot of baseline eGFR by the WGS and strata of EA is presented in **Figure 1A**. On visual inspection of this data, the effect of the WGS on eGFR appeared to be consistent across strata of EA, hence, we anticipated that the interaction term between the WGS and EA in our models would not be significant. In unadjusted models (*models 1-2*), both the WGS and EA were independently associated with eGFR (**Table 2** Results of interaction analysis). A one-sd increase in the WGS was associated with 1.61 mL/min/1.73m^2 lower eGFR (*model 1*, B ± se = -1.61 ± 0.28, p=1.5 x10^{-8}), while those with low EA were observed to have the lowest mean eGFR (*model 1*, low vs high EA, B ± se = -8.74 ± 0.67, p=5.9 x10^{-38}, **Table 2**). Addition of an interaction term (WGS x EA) did not contribute to the model (*model 2 vs model 1*, $\Delta R^2_{adjusted}$ = -0.0001; $F_{(2,\ 3360)}$ = 0.664, p=0.512). Adjustment for covariates (*models 3-4*; age, age^2, sex, and the first 10 PCs) did not affect the association of the WGS with baseline eGFR. However, the association between EA and baseline eGFR disappeared due to strong confounding by age.

Longitudinal analysis

Median follow-up duration was 11 years (interquartile range: 4.6 – 11.9 years). In the total population, the average change in eGFR was -0.927 mL/min/1.73m^2 per year (sd = 0.385). A plot of eGFR change by the WGS and strata of EA is presented in **Figure 1B**. In this figure, the WGS is shown to have its strongest effect on eGFR change in those with low EA. In those with medium or high EA, the WGS had no apparent added effect on eGFR change. A trend in mean eGFR change was observed across EA levels, with those with lower EA having increasingly faster rates of decline on average.

In unadjusted models (*models 1-2*), a one-sd increase in the WGS was associated with 0.016 ml/min/m^2 per year faster eGFR decline (*model 1*, B ± se = -0.016 ± 0.007, p = 0.014, **Table 2**) and EA (*model 1*, low vs high EA, B ± se = -0.125 ± 0.016, p = 3.3 x10^{-15}) was also independently associated with rate of kidney function decline. Adjustment for covariates (*models 3-4*; age, age^2, sex, and the first 10 PCs) increased the effect of the WGS on eGFR change (*model 3*, B ± se = -0.027 ± 0.006, p=2.3 x10^{-5}), while attenuating the effect of EA on eGFR change (*model 3*, low vs high EA, B ± se = -0.054 ± 0.016, p = 7.9 x10^{-4}). A WGS x EA interaction term was in

Figure 1. Plots of eGFR versus weighted genetic score for reduced eGFR, by educational attainment. Upper panels (A) show cross-sectional eGFR (mL/min/1.73m²), and lower panels (B) show annual change in eGFR (mL/min/1.73m² per year), stratified by educational attainment (high, medium, low). Regression lines with 95% confidence interval are derived from unadjusted ordinary linear regression.



the expected direction (*model 4*, low vs high EA, B ± se = -0.036 ± 0.015, p=0.017), suggesting that the joint effect of the WGS and EA is greater than the sum of their main effects. The contribution of the overall interaction term between the WGS and EA was modest (*model 4 vs model 3*, $\Delta R^2_{adjusted}$ = 0.0012) but significant ($F_{(2, 3327)}$ =3.32, p=0.036).

Table 2. Results of interaction analyses

A. eGFR (mL/min/1.73m²)

	Model 1			***Model 2***			***Model 3***			***Model 4***			***Model 5***			***Model 6***		
	$R^2_{adjusted}$ =0.063			$R^2_{adjusted}$ =0.063			$R^2_{adjusted}$ =0.446			$R^2_{adjusted}$ =0.445			$R^2_{adjusted}$ =0.459			$R^2_{adjusted}$ =0.459		
	B	**se**	**p**	**B**	**se**	**p**	**B**	**se**	**p**	**B**	**se**	**p**	**B**	**se**	**p**	**B**	**se**	**p**
Intercept	99.27	0.52	**0.00**	99.27	0.52	**0.00**	91.39	0.56	**0.00**	91.39	0.56	**0.00**	93.56	2.42	0.00	93.62	2.43	**0.00**
WGS (per sd)	-1.61	0.28	**1.5 x10^{-8}**	-2.04	0.55	**1.9 x10^{-4}**	-1.76	0.22	**1.4 x10^{-15}**	-2.12	0.42	**5.3 x10^{-7}**	-1.71	0.22	**1.1 x10^{-14}**	-1.93	0.42	**4.7 x10^{-6}**
low EA	-8.74	0.67	**5.9 x10^{-38}**	-8.74	0.67	**5.7 x10^{-38}**	0.24	0.56	0.674	0.23	0.56	0.677	1.37	0.58	**0.018**	1.37	0.58	**0.018**
medium EA	-2.18	0.77	**4.9 x10^{-3}**	-2.18	0.77	**5.0 x10^{-3}**	0.06	0.60	0.914	0.07	0.60	0.910	0.61	0.61	0.314	0.61	0.61	0.315
high EA	ref			ref			ref			ref			ref			ref		
WGS * low EA	-	-	-	0.77	0.69	0.265	-	-	-	0.60	0.53	0.256	-	-	-	0.45	0.53	0.395
WGS * medium EA	-	-	-	0.29	0.77	0.711	-	-	-	0.29	0.60	0.628	-	-	-	0.05	0.60	0.936
WGS * high EA	-	-	-	ref			-	-	-	ref			-	-	-	ref		

Table 2. *(continued)*. Results of interaction analyses

B. annual eGFR change (mL/min/1.73m² per year)

	Model 1			***Model 2***			***Model 3***			***Model 4***			***Model 5***			***Model 6***		
	$R^2_{adjusted}$ =0.040			$R^2_{adjusted}$ =0.041			$R^2_{adjusted}$ =0.108			$R^2_{adjusted}$ =0.109			$R^2_{adjusted}$ =0.124			$R^2_{adjusted}$ =0.125		
	B	**se**	**p**	**B**	**se**	**p**	**B**	**se**	**p**	**B**	**se**	**p**	**B**	**se**	**p**	**B**	**se**	**p**
Intercept	-1.089	0.041	**0.00**	-1.090	0.041	**0.00**	-0.695	0.048	**0.00**	-0.697	0.048	**0.00**	-0.341	0.085	**0.00**	-0.346	0.085	**0.00**
WGS (per sd)	-0.016	0.007	**0.014**	0.004	0.013	0.746	-0.027	0.006	**2.3×10^{-5}**	-0.008	0.012	0.520	-0.026	0.006	**3.7×10^{-5}**	-0.008	0.012	0.488
low EA	-0.125	0.016	**3.3×10^{-15}**	-0.124	0.016	**3.7×10^{-15}**	-0.054	0.016	**7.9×10^{-4}**	-0.054	0.016	**8.1×10^{-4}**	-0.056	0.017	**8.5×10^{-4}**	-0.056	0.017	**8.4×10^{-4}**
medium EA	-0.042	0.018	**0.018**	-0.042	0.018	**0.018**	-0.026	0.017	0.131	-0.026	0.017	0.130	-0.022	0.017	0.213	-0.022	0.017	0.211
high EA	ref			ref			ref			ref			ref			ref		
WGS * low EA	-	-	-	-0.037	0.016	**0.018**	-	-	-	-0.036	0.015	**0.017**	-	-	-	-0.034	0.015	**0.027**
WGS * medium EA	-	-	-	-0.011	0.018	0.537	-	-	-	-0.009	0.017	0.588	-	-	-	-0.010	0.017	0.572
WGS * high EA	-	-	-	ref			-	-	-	ref			-	-	-	ref		

Results of interaction analyses.
Model 1: WGS + EA
Model 2: *model 1* + WGS x EA
Model 3: WGS + EA + age + age² + sex + PCs 1-10
Model 4: *model 3* + WGS x EA
Model 5: WGS + EA + BMI + SBP + glucose + cholesterol + smoking + age + age² + sex + PCs 1-10
Model 6: *model 5* + WGS x EA
For longitudinal analysis, baseline eGFR was added to each model. Bold p-values indicate significance at the =0.05 level.
Abbreviations: eGFR, estimated glomerular filtration rate; WGS, weighted genetic score; EA, educational attainment; BMI, body-mass index; SBP, systolic blood pressure; PCs, genetic principal components. Coefficients for covariates are omitted for clarity.

The effects of potential mediators (i.e. BMI, SBP, glucose, total cholesterol, and smoking status) of the interaction were assessed in our final models (*model 5-6*). Addition of these risk factors did not affect the association between the WGS and eGFR change (*model 5*, B ± se = -0.027 ± 0.006, p =2.32 $x10^{-5}$) whereas the effect of EA was slightly attenuated (*model 5*, low vs high EA, B ± se = -0.047 ± 0.016, p=4.33 $x10^{-3}$), suggesting potential mediation by these risk factors. Potential mediation was further supported by the finding that the overall interaction effect was only borderline significant after addition of these risk factors (*model 6 vs model 5*, $\Delta R^2_{adjusted}$ = 0.0010; $F_{(2,\ 3213)}$ = 2.78, p=0.062), although the interaction effect of the WGS with low vs high EA was not attenuated and remained nominally significant (model 6, B ± se = -0.034 ± 0.015, p=0.027).

Sensitivity analysis

The WGS did not show significantly different distributions between categories of EA. However, **Figure 1** and **Supplementary Figure S1** are suggestive of slight overrepresentation of a higher WGS in those with lower EA and a lower WGS in those with higher EA. To minimize bias due to potentially influential observations, we excluded eight observations that exceeded a more stringent cut-off of three sds from the mean. These sensitivity analyses yielded essentially the same results as our main analyses, although significance decreased slightly due to reduced statistical power (data not shown).

Furthermore, we repeated all analyses for eGFR estimated from serum creatinine only (eGFRcrea), and eGFR estimated from serum cystatin C only (eGFRcysc). Results were generally consistent with our main analysis, with EA being more strongly associated with eGFRcysc than with eGFRcrea. Similarly, interaction effects between the WGS and EA were more pronounced for eGFRcysc than for eGFRcrea (data not shown).

Finally, we repeated the interaction analyses using LMM only. Here, despite some minor discrepancy with longitudinal estimates from OLS, effect estimates were generally and directionally consistent with the OLS analysis (**Supplementary Table S1**), and a three-way interaction term to assess the modifying effect of EA on WGS on eGFR change (WGS x EA x time) was again significant (**Supplementary Table S2**).

DISCUSSION

In the present study, we investigated the effects of genetic factors (summarized by a weighted genetic score, WGS) and educational attainment (EA), as well as the interaction between the WGS and EA, on kidney function outcomes. We observed additive effects of the WGS and EA for baseline eGFR in cross-sectional analyses, although these were not robust to covariate adjustment. In longitudinal analyses, low EA interacted with high WGS, resulting in faster eGFR decline. This amplifying effect of low EA on genetic risk could not entirely be explained by a less favorable renal risk factor profile in those with low EA (i.e. higher BMI, higher SBP, higher glucose, higher cholesterol, and higher prevalence of smoking).

In the present study, participants with low EA had similar genetic risk of CKD compared to those with higher EA, since the WGS was equally distributed to each stratum of EA. However, the impact of genetic risk on annual eGFR decline was observed to be larger in those with low EA, resulting in a disproportionally high risk of CKD for the most vulnerable in terms of EA and genetic predisposition. Low EA is unlikely to directly amplify genetic risk of CKD. Rather, it may act through a range of interrelated downstream effects of low EA such as lower income, poor health behavior, poor health care access, and higher prevalence of traditional renal risk factors[8,9]. In our analyses, the interaction effect was only partly explained by traditional renal risk factors. Therefore, other factors likely exist that explain the interaction between EA and CKD. These may include factors with socioeconomic gradients such as health literacy[29], occupational exposures and infections[30], whose influence may not be captured by traditional risk factors.

Individually, the 63 SNPs that were identified in previous GWAS on eGFRcrea[14] have small effects. The WGS aggregates these effects, thereby greatly increasing statistical power compared to using single SNP effects. Therefore, the WGS is a practical summary score of genetic risk for reduced kidney function. However, some limitations with regards to the WGS must be addressed. The WGS only explained a small fraction of between-individual variation in eGFR in PREVEND. In addition, participants with an equal WGS may have different underlying risk variants. With ever-increasing sample sizes for GWAS, it is expected that larger numbers of SNPs can be detected with greater precision, thereby resulting in a WGS that is a more comprehensive summary measure of genetic risk. Furthermore, by using a WGS in interaction analysis, it is implicitly assumed that all genetic

variants included in the WGS have directionally consistent interaction effects with EA. Another implicit assumption is that the same set of genetic variants affect eGFR in each category of EA. To check these assumptions, single SNP interaction effects would need to be assessed, but this requires infeasibly large sample sizes and is therefore beyond the scope of the present study. Future research may include genome-wide interaction studies to identify the specific genetic variants whose effects are modified by EA. Similar studies have been done for blood pressure, BMI and lipids for specific exposures such as smoking, alcohol use and physical activity[31-34].

For the longitudinal analyses, we chose to report results from a two-step method in which we used individual eGFR trajectories modelled with LMM as outcome variable in OLS regression. This allows for straightforward estimation of model R^2 and intuitive interpretation of the WGS x EA two-way interaction term. The two-step approach potentially comes at the cost of introducing false precision in eGFR trajectories given that random variation in eGFR measurements during follow-up is ignored to an extent. This may explain that in previous study in PREVEND, a WGS comprising 63 SNPs showed similar effects on eGFR change compared to the present study, but did not reach statistical significance in LMM analysis[17]. Alternatively, the effects of the WGS, EA, and the WGS x EA interaction term on eGFR change can also be modelled in a single LMM model, taking into account the random variation and correlation between eGFR measurements. However, R^2 estimation is not straightforward in LMM models, and the effect of the interaction on eGFR change requires modelling a three-way interaction term (WGS x EA x time), the interpretation of which is less intuitive compared to that of a two-way interaction term. We performed sensitivity analyses using an LMM model only. Notwithstanding some discrepancies with the OLS analysis regarding effect size and statistical significance, the results from LMM were directionally consistent with OLS analysis and therefore our conclusions remain unchanged.

Given that in the present study, the interaction between the WGS and EA resulted in accelerated rates of eGFR decline, we hypothesize that this interaction also results in increased rates of CKD. However, given the large sample size needed to find significant interaction effects on categorical outcomes, we opted not to perform analyses with incident CKD as outcome. Further research in larger samples

is needed to assess whether the interaction indeed leads to an increase in CKD incidence, with a definition of CKD based on clinically relevant cut-off values.

Our study adds to the literature on socioeconomic disparities in CKD as it is the first to present evidence of gene-environment interaction between a WGS, based on SNPs associated with eGFR, and EA. Major strengths of this study include the availability of multiple eGFR estimates per individual, that are based on both serum creatinine and cystatin C values, that were measured in one run allowing precise estimation of glomerular filtration rate, and the considerable follow-up duration. Several limitations, other than those already discussed, need to be addressed. First, the present study population consists of participants of European ancestry exclusively, sampled from a relatively high-income population (i.e. the population of Groningen, the Netherlands). Therefore, the generalizability of these findings to non-European, lower-income populations may be limited. Second, the interaction effects of genetic risk and EA on rate of kidney function decline that we found are modest and therefore require replication in independent samples. Third, the observational nature of this study precludes causal conclusions. Finally, a higher attrition rate was observed in those with low education. This may have resulted in bias towards the null, or underestimation of effects, due to reduced power and precision of kidney decline outcomes in this group.

Knowledge of the interaction that we found in our longitudinal analyses is unlikely to be useful for risk stratification for preventive medicine, due to the rather modest effects. However, our results may inform public health policy as they provide insights into the mechanisms that underlie socioeconomic disparities in CKD. For example, it is possible that downstream effects of low EA contribute to an environment that activates genetic pathways that are detrimental for kidney health. Conversely, deleterious genetic effects are suggested to be completely mitigated by high EA and its downstream effects, at least with regards to kidney function decline. Future study is needed to identify which factors are responsible for this modifying effect, as these factors are potential targets for intervention to reduce socioeconomic disparities in CKD.

In conclusion, our findings provide population level insights on the mechanisms underlying socioeconomic disparities in CKD. We observed that a WGS, as a summary measure of genetic risk, and EA have independent effects on the rate of kidney function decline. Furthermore, our results suggest a subtle amplifying effect of low EA on genetic risk of eGFR. Traditional kidney risk factors that are known downstream effects of low EA (i.e. higher BMI, higher SBP, higher glucose, higher cholesterol, and higher prevalence of smoking) did not explain the amplifying effect on the WGS, which warrants further investigation.

REFERENCES

1. Levey AS, de Jong PE, Coresh J, et al. The definition, classification, and prognosis of chronic kidney disease: A KDIGO controversies conference report. *Kidney Int*. 2011;80(1):17-28.
2. Gansevoort RT, Correa-Rotter R, Hemmelgarn BR, et al. Chronic kidney disease and cardiovascular risk: Epidemiology, mechanisms, and prevention. *The Lancet*. 2013;382(9889):339-352.
3. Stenvinkel P. Chronic kidney disease: A public health priority and harbinger of premature cardiovascular disease. *J Intern Med*. 2010;268(5):456-467.
4. Jha V, Garcia-Garcia G, Iseki K, et al. Chronic kidney disease: Global dimension and perspectives. *The Lancet*. 2013;382(9888):260-272.
5. Grams ME, Chow EKH, Segev DL, Coresh J. Lifetime incidence of CKD stages 3-5 in the United states. *Am J Kidney Dis*. 2013;62(2):245-252.
6. Vart P, Gansevoort RT, Joosten MM, Bultmann U, Reijneveld SA. Socioeconomic disparities in chronic kidney disease: A systematic review and meta-analysis. *Am J Prev Med*. 2015;48(5):580-592.
7. Zeng X, Liu J, Tao S, Hong HG, Li Y, Fu P. Associations between socioeconomic status and chronic kidney disease: A meta-analysis. *J Epidemiol Community Health*. 2018;72(4):270-279.
8. Vart P, Gansevoort RT, Crews DC, Reijneveld SA, Bultmann U. Mediators of the association between low socioeconomic status and chronic kidney disease in the United states. *Am J Epidemiol*. 2015;181(6):385-396.
9. Thio CH, Vart P, Kieneker LM, Snieder H, Gansevoort RT, Bültmann U. Educational level and risk of chronic kidney disease: Longitudinal data from the PREVEND study. *Nephrology Dialysis Transplantation*. 2018.
10. Lei HH, Perneger TV, Klag MJ, Whelton PK, Coresh J. Familial aggregation of renal disease in a population-based case-control study. *J Am Soc Nephrol*. 1998;9(7):1270-1276.
11. Satko SG, Freedman BI. The familial clustering of renal disease and related phenotypes. *Med Clin North Am*. 2005;89(3):447-456.
12. Skrunes R, Svarstad E, Reisaeter AV, Vikse BE. Familial clustering of ESRD in the Norwegian population. *Clin J Am Soc Nephrol*. 2014;9(10):1692-1700.
13. Arpegard J, Viktorin A, Chang Z, de Faire U, Magnusson PK, Svensson P. Comparison of heritability of cystatin C- and creatinine-based estimates of kidney function and their relation to heritability of cardiovascular disease. *J Am Heart Assoc*. 2015;4(1):e001467.
14. Gorski M, van der Most PJ, Teumer A, et al. 1000 genomes-based meta-analysis identifies 10 novel loci for kidney function. *Scientific Reports*. 2017;7:45040.
15. O'Seaghdha CM, Yang Q, Wu H, Hwang SJ, Fox CS. Performance of a genetic risk score for CKD stage 3 in the general population. *Am J Kidney Dis*. 2012;59(1):19-24.
16. Ma J, Yang Q, Hwang S, Fox CS, Chu AY. Genetic risk score and risk of stage 3 chronic kidney disease. *BMC nephrology*. 2017;18(1):32.
17. Thio CH, van der Most, Peter J, Nolte IM, et al. Evaluation of a genetic risk score based on creatinine-estimated glomerular filtration rate and its association with kidney outcomes. *Nephrol Dial Transplant*. 2017;33(10):1757-64.
18. Liu SY, Walter S, Marden J, et al. Genetic vulnerability to diabetes and obesity: Does education offset the risk? *Soc Sci Med*. 2015;127:150-158.
19. van Zon SKR, Reijneveld SA, van der Most PJ, Swertz MA, Bultmann U, Snieder H. The interaction of genetic predisposition and socioeconomic position with type 2 diabetes mellitus: Cross-sectional and longitudinal analyses from the Lifelines cohort and Biobank study. *Psychosom Med*. 2018;80(3):252-262.
20. Amin V, Böckerman P, Viinikainen J, et al. Gene-environment interactions between education and body mass: Evidence from the UK and Finland. *Soc Sci Med*. 2017;195:12-16.
21. Barcellos SH, Carvalho LS, Turley P. Education can reduce health differences related to genetic risk of obesity. *Proc Natl Acad Sci U S A*. 2018;115(42):E9765-E9772.

22. Pinto-Sietsma SJ, Janssen WM, Hillege HL, Navis G, De Zeeuw D, De Jong PE. Urinary albumin excretion is associated with renal functional abnormalities in a nondiabetic population. *J Am Soc Nephrol*. 2000;11(10):1882-1888.

23. Grubb A, Blirup-Jensen S, Lindstrom V, et al. First certified reference material for cystatin C in human serum ERM-DA471/IFCC. *Clin Chem Lab Med*. 2010;48(11):1619-1621.

24. Inker LA, Schmid CH, Tighiouart H, et al. Estimating glomerular filtration rate from serum creatinine and cystatin C. *N Engl J Med*. 2012;367(1):20-29.

25. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS genetics*. 2006;2(12):e190.

26. UNESCO Institute for Statistics. International standard classification of education. 2011.

27. R Core Team. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria.

28. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*. 2014.

29. Magnani JW, Mujahid MS, Aronow HD, et al. Health literacy and cardiovascular disease: Fundamental relevance to primary and secondary prevention: A scientific statement from the Aamerican Heart Association. *Circulation*. 2018;138(2):e48-e74.

30. Obrador GT, Schultheiss UT, Kretzler M, et al. Genetic and environmental risk factors for chronic kidney disease. *Kidney Int Supp*. 2017;7(2):88-106.

31. Justice AE, Winkler TW, Feitosa MF, et al. Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits. *Nat Commun*. 2017;8:14977.

32. Sung YJ, de Las Fuentes L, Winkler TW, et al. A multi-ancestry genome-wide study incorporating gene–smoking interactions identifies multiple new loci for pulse pressure and mean arterial pressure. *Hum Mol Genet*. 2019;28(15):2615-2633.

33. de Vries PS, Brown MR, Bentley AR, et al. Multiancestry genome-wide association study of lipid levels incorporating gene-alcohol interactions. *Am J Epidemiol*. 2019;188(6):1033-1054.

34. Kilpeläinen TO, Bentley AR, Noordam R, et al. Multi-ancestry study of blood lipid levels identifies four loci interacting with physical activity. *Nat Commun*. 2019;10(1):376.

Supplementary Material

CHAPTER 7

Table S1. Results of interaction analyses using linear mixed models

eGFR (mL/min/1.73m²)																		
	Model 1			***Model 2***			***Model 3***			***Model 4***			***Model 5***			***Model 6***		
	B	**se**	**p**	**B**	**se**	**p**	**B**	**se**	**p**	**B**	**se**	**p**	**B**	**se**	**p**	**B**	**se**	**p**
Intercept	99.331	0.517	**0.000**	99.330	0.517	**0.000**	90.950	0.519	**0.000**	90.960	0.519	**0.000**	95.910	2.296	**0.000**	95.950	2.299	**0.000**
WGS (per sd)	-1.633	0.280	**0.000**	-1.893	0.536	**0.000**	-1.801	0.215	**0.000**	-2.097	0.411	**0.000**	-1.766	0.217	**0.000**	-1.945	0.412	**0.000**
low EA	-8.620	0.661	**0.000**	-8.626	0.662	**0.000**	0.324	0.550	0.555	0.319	0.550	0.562	1.375	0.566	**0.015**	1.370	0.566	**0.016**
medium EA	-2.074	0.764	**0.007**	-2.079	0.764	**0.007**	0.182	0.590	0.758	0.181	0.590	0.759	0.680	0.596	0.254	0.675	0.596	0.257
high EA	ref			ref			ref			ref			ref			ref		
time	-0.787	0.032	**0.000**	-0.787	0.032	**0.000**	0.096	0.037	**0.009**	0.096	0.037	**0.009**	0.088	0.038	**0.021**	0.088	0.038	**0.016**
WGS * low EA	-	-	-	0.574	0.676	0.396	-	-	-	0.544	0.520	0.295	-	-	-	0.415	0.523	0.427
WGS * medium EA	-	-	-	-0.127	0.048	**0.008**	-	-	-	0.183	0.584	0.754	-	-	-	-0.025	0.586	0.965
WGS * high EA	-	-	-	ref			-	-	-	ref			-	-	-	ref		
WGS * time	-0.026	0.018	0.149	0.021	0.033	0.524	-0.031	0.018	0.084	0.019	0.032	0.554	-0.033	0.018	0.063	0.015	0.032	0.637
low EA * time	-0.283	0.042	**0.000**	-0.281	0.042	**0.000**	-0.144	0.043	**0.001**	-0.140	0.043	**0.001**	-0.147	0.043	**0.001**	-0.144	0.043	**0.001**
medium EA * time	-0.129	0.048	**0.007**	-0.127	0.048	**0.008**	-0.093	0.005	**0.047**	-0.091	0.047	0.052	-0.081	0.047	0.083	-0.079	0.047	0.090
high EA * time	ref			ref			ref			ref			ref					
WGS * low EA * time	-	-	-	-0.104	0.043	**0.015**	-	-	-	-0.114	0.042	**0.007**	-	-	-	-0.116	0.042	**0.006**
WGS * medium EA * time	-	-	-	-0.007	0.048	0.885	-	-	-	0.000	0.047	0.998	-	-	-	0.004	0.047	0.931
WGS * high EA * time	-	-	-	ref			-	-	-	ref			-	-	-	ref		

Results of interaction analyses using linear mixed models. Model parameters are estimated using restricted maximum likelihood.

Model 1: (WGS + EA) x time + random(intercept + time)
Model 2: *model 1* + (WGS x EA) x time
Model 3: (WGS + EA) x time + age + age² + sex + PCs 1-10 + random(intercept + time)
Model 4: *model 3* + (WGS x EA) x time
Model 5: (WGS + EA) x time + BMI + SBP + glucose + cholesterol + smoking + age + age² + sex + PCs 1-10 + random(intercept + time)
Model 6: *model 5* + (WGS x EA) x time
Bold p-values indicate significance at the =0.05 level.
Abbreviations: eGFR, estimated glomerular filtration rate; WGS, weighted genetic score; EA, educational attainment; BMI, body-mass index; SBP, systolic blood pressure; PCs, genetic principal components.Coefficients for covariates are omitted for clarity.

Table S2. LMM model comparisons

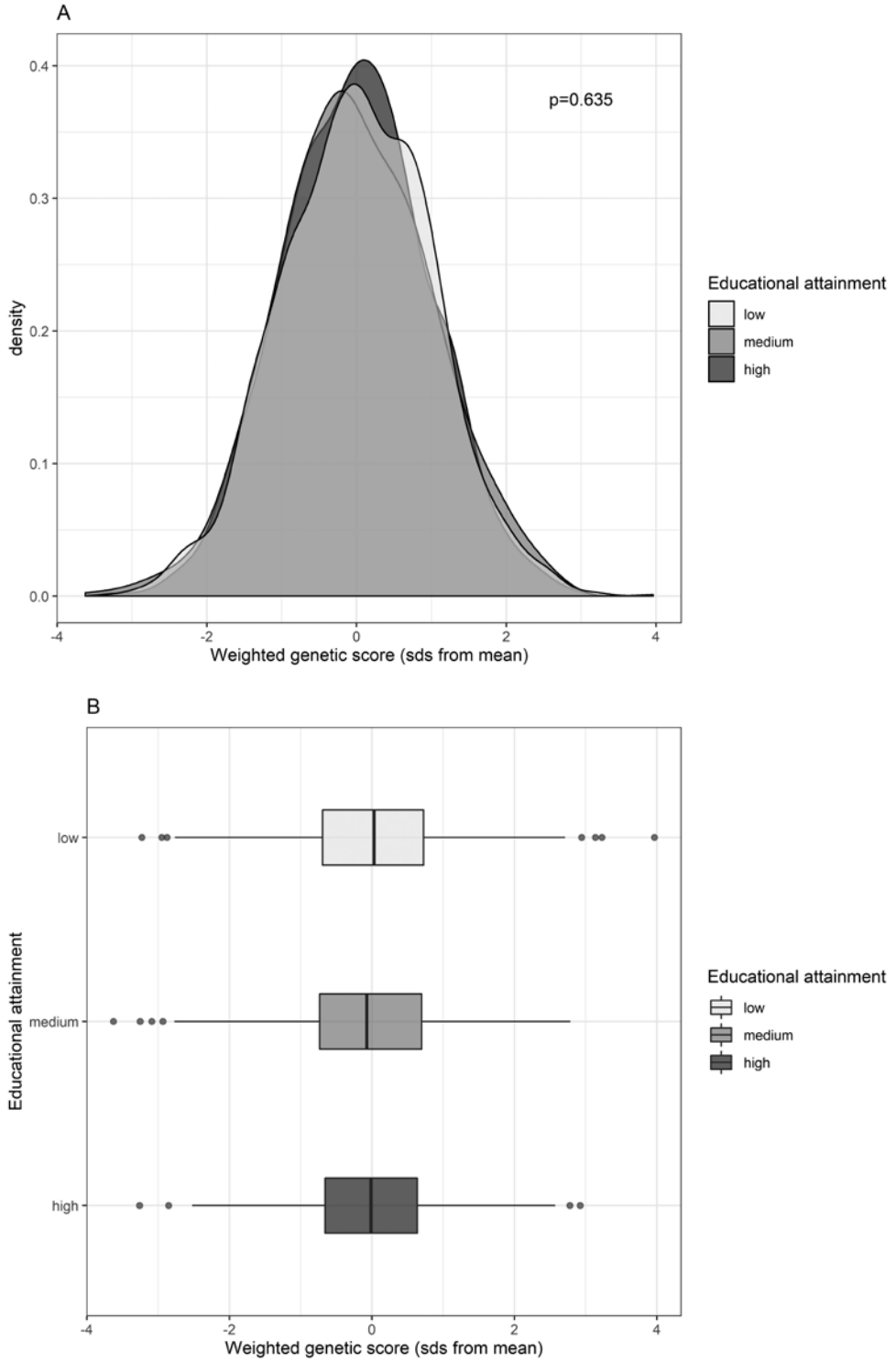
Model	df	AIC	BIC	logLik	deviance	χ^2	Δdf	p
Model 1	12	87311	87400	-43644	87287			
Model 2	16	87311	87430	-43640	87279	8.1169	4	0.087
Model 3	25	85221	85406	-42585	85171			
Model 4	29	85218	85433	-42580	85160	10.306	4	0.036
Model 5	31	82072	82301	-41005	82010			
Model 6	35	82070	82328	-41000	82000	10.698	4	0.030

Comparison of nested models with and without an interaction term for WGS x EA. Models were refitted from restricted maximum likelihood to maximum likelihood.

Model 1: (WGS + EA) x time + random(intercept + time)
Model 2: *model 1* + (WGS x EA) x time
Model 3: (WGS + EA) x time + age + age^2 + sex + PCs 1-10 + random(intercept + time)
Model 4: *model 3* + (WGS x EA) x time
Model 5: (WGS + EA) x time + BMI + SBP + glucose + cholesterol + smoking + age + age^2 + sex + PCs 1-10 + random(intercept + time)
Model 6: *model 5* + (WGS x EA) x time

Df, degrees of freedom; AIC, Akaike information criterion; BIC, Bayesian information criterion; logLik, log likelihood.

Supplementary Figure 1



7

Investigating causal effects of educational attainment on kidney outcomes: a Mendelian randomization study

Chris HL Thio, Peter J van der Most, Ute Bültmann, Ron T Gansevoort and Harold Snieder

In preparation

CHAPTER

8

ABSTRACT

Introduction. Educational attainment (EA) is associated with reduced risk of chronic kidney disease (CKD), higher kidney function (estimated glomerular filtration rate, eGFR) and less kidney damage (urinary albumin-to-creatinine ratio, UACR). We aimed to ascertain whether these associations constitute causal relations.

Methods. Using a two-sample Mendelian randomization (MR) design, we used 1271 single nucleotide polymorphisms associated with years of schooling to genetically predict EA (gEA), thereby minimizing confounding. We used genome-wide association study summary data for a number of kidney traits from up to 567,460 participants of European descent.

Results. Effects of gEA are per one SD (4.2 years). Higher gEA was associated with higher cystatin C-estimated GFR (B=3.2%, 95%CI 1.9% to 4.6%, $p=2.4x10^{-6}$), but not with creatinine-estimated GFR. Contrary to expectations, higher gEA was associated with higher inverse normally transformed UACR (B=0.06, 95%CI: 0.043 to 0.076, $p=2.5x10^{-12}$). Higher gEA was associated with lower urinary creatinine concentration ($p=1.2x10^{-60}$), leading us to hypothesize confounding by creatinine metabolism (e.g. muscle mass) explains the positive association with UACR. However, in 24-hour urinary data from the Lifelines Cohort (N=12,675), we found no effect on 24h creatinine excretion (p=0.861), dismissing muscle mass as a confounding factor. Instead, higher gEA was associated with higher 24-h urinary albumin excretion (p=0.019), suggesting higher EA indeed increases kidney damage.

Conclusion. In this MR study, we found inconsistent effects of gEA on eGFR and even a deleterious effect of gEA on albuminuria. The results of this study warrant further investigation, and plead against a causal protective effect of EA on CKD.

Keywords: Chronic kidney disease, Educational attainment, Mendelian randomization

INTRODUCTION

Chronic kidney disease (CKD) is a heterogeneous group of disorders defined by sustained reduced kidney function and signs of kidney damage[1,2]. It is a risk factor for cardiovascular morbidity and mortality[3,4] and it may progress to end-stage renal disease. The global prevalence of CKD is 10-15%, and the management of CKD and its consequences poses a heavy burden on patients and health care resources.

Socioeconomic gradients in CKD rates are observed: indicators of socioeconomic status such as higher educational attainment (EA) are associated with increased kidney function and reduced kidney damage in traditional observational studies[5-7]. It has been suggested that EA affects CKD risk through a number of intermediate factors observed to be less prevalent in those with higher EA, such as smoking, hypertension, and diabetes[8,9]. However, this proposition is predicated on the assumption that there is a protective causal effect of EA on CKD. Whether such a causal effect exists remains uncertain due to potential unobserved confounding and reverse causation in traditional observational studies.

A Mendelian randomization (MR) study may help in evaluating causality using observational data, when experiments are impractical or undesirable. This method utilizes genetic variants as proxies for exposures such as EA in instrumental variable analysis[10-12]. Due to the random assignment of genetic variants during meiosis, these variants are independent of confounding factors. Furthermore, given that genetic variants are fixed throughout life, reverse causation is unlikely. Previous MR studies established protective effects of EA on coronary heart disease[13], and intermediates for coronary heart disease, such as lipid levels, BMI, blood pressure, and smoking behavior[13-16]. Given the purported overlap in pathophysiology and risk factors between cardiovascular and renal disease[17], higher EA is expected to also be protective of CKD.

Thus, using MR, we aimed to ascertain whether EA has protective causal effects on kidney function and kidney damage.

METHODS

Overall design

We applied a two-sample MR study design[11,12,18], which utilizes single nucleotide polymorphisms (SNPs) as instrumental variables to minimize confounding, requiring only summary level statistics from large-scale genome-wide association studies (GWAS). The GWAS from which summary data were leveraged are listed in **Table 1**. In secondary analyses in the Lifelines Cohort study, we applied one-sample, individual level MR to secondary kidney outcomes to assess the validity of our findings. All analyses were performed in populations of European descent.

Outcome definitions

For our main two-sample MR analyses, kidney outcomes were defined as described in their original GWAS studies[19,20]. Briefly, kidney function was approximated by glomerular filtration rate, estimated by CKD-EPI equations for creatinine[21] (eGFRcrea) and cystatin C[22] (eGFRcysc), transformed to their natural logarithm, and regressed on age and sex. The resulting unstandardized residuals of *ln*eGFRcrea and *ln*eGFRcysc were then used as outcome variables. Kidney damage was approximated by *ln*-transformed urinary albumin-to-creatinine ratio (*ln*UACR), residualized to age and sex, and then inverse normally transformed.

Educational attainment

In the original GWAS on EA[23], data were restricted to European-ancestry individuals that passed the cohort's quality control and whose EA was measured at an age of at least 30 years. EA was constructed by mapping each major educational qualification that can be identified from the cohort's survey measure to an International Standard Classification of Education (ISCED) 1997[24] category and imputing a years-of-schooling equivalent for each ISCED category (**Supplementary Table S1**). The EA phenotype was then standardized; each SD represents 4.2 years of schooling.

Two-sample Mendelian randomization

We used 1271 independent SNPs (linkage disequilibrium, LD: $r^2<0.1$) associated with years of schooling[23] in populations of European ancestry as instrumental variables to genetically predict EA (gEA). The effects of these 1271 SNPs were extracted from European ancestry GWAS on eGFRcrea[19], eGFRcysc[20], and UACR (Teumer et al, Nature Communications, 2019, in press). We harmonized

datasets on SNP alleles, and removed palindromic SNPs with intermediate allele frequencies (minor allele frequency >0.42). MR analyses were performed using the *TwoSampleMR* R-package[25] in R software version 3.4.2[26]. As main analysis, we performed an inverse variance weighted (IVW) meta-analysis of SNP effects. As sensitivity analyses, we performed weighted median analysis[27], MR Egger regression[28], and mode-based MR[29] to test robustness to varying degrees of violations of MR assumptions, in particular those due to pleiotropy. Leave-one-out analyses were performed to detect disproportionately influential SNPs. In an additional sensitivity analyses, we pruned the data by clumping SNPs with a more stringent LD cut-off of r^2<0.001 to ensure independence of SNPs. In this pruned dataset, we repeated all MR analyses and, in addition, performed Mendelian Randomization Pleiotropy Residual Sum and Outlier (MR-PRESSO) analysis to detect heterogeneity and account for outlying, and therefore potentially pleiotropic, SNP effects using the *MR-PRESSO* R package[30]. The MR Steiger test[31] was performed to infer directionality of effects.

Secondary analyses

Two-sample Mendelian Randomization on urinary creatinine concentration

In our main analyses, we found an unexpected detrimental effect of gEA on UACR. In secondary analyses, we aimed to ascertain whether this was due to the creatinine component in UACR by examining the effect of gEA on urinary creatinine concentrations (UcreaC) in the UK Biobank (UKBB, www.nealelab.is/uk-biobank/) using the two-sample MR methods described above.

2-Stage least squares analysis in Lifelines on creatinine and albumin excretion

In individual participant data from the Lifelines Cohort Study, we examined the gEA-UACR relation. We used data of unrelated, genotyped participants. The Lifelines Cohort Study is a multidisciplinary prospective population-based cohort study with a unique 3-generation design that examines health and health-related behavior of 165,729 participants living in the north-eastern region of the Netherlands (www.lifelines.nl/researcher). Participants were recruited through their general practitioner or through participating family members. Additionally, there was the option to self-register. Details on study design, participant selection and genotyping in Lifelines have been described previously[32-35]. Genotyped participants aged ≥30 years with completed questionnaire data on educational attainment were included for this analysis (N=12,675). The Lifelines Cohort Study

was conducted according to the guidelines in the Declaration of Helsinki, and all procedures involving human subjects were approved by the Medical Ethics Committee of the University Medical Center Groningen. Written informed consent was obtained from all participants during the visit to one of the research centers.

We performed analyses on UACR, and additionally, urinary concentrations and 24h excretions of creatinine (UcreaC and UcreaE) and albumin (UAC and UAE) separately, as well as urinary volume obtained from 24h urine collections. For all secondary analyses, 24h urinary albumin concentration values were left-truncated at the limit of detection (LOD, 2.3 mg/L). In case variables were right-skewed, we applied a *ln*-transformation. After transformation, we removed outliers deviating >4 standard deviations from the mean. The normalized variables were then regressed on age, age^2, age^3, sex and the first ten genetically derived principal components (to account for population structure) to obtain unstandardized residuals. These residuals were subsequently used as dependent variables as described below. For UACR, we in addition inverse normally transformed the residuals (*int ln*UACR).

Highest educational qualification was assessed through self-report questionnaires, mapped to the ISCED, and then converted into years of schooling. We constructed a 1271-SNP weighted genetic score (WGS) for years of schooling, weighted for effects originally reported by Lee et al,[23] and used this WGS as instrumental variable to genetically predict EA. We performed two-stage least squares (2SLS) analyses using R version 3.4.2. In the first stage, we regressed years of schooling on the WGS. The model-predicted estimate of years of schooling (gEA), resulting from this first stage regression, is an unbiased genetic proxy of EA due to the assumed random assortment of the WGS. The gEA was then used as independent variable in second stage regression. gEA was divided by 4.2 years to allow comparison with two-sample MR results. Dependent variables in this second stage were the previously mentioned unstandardized residuals of urinary outcomes.

Genetic correlations

To ascertain whether MR estimates were consistent with genetic correlations between EA and the kidney outcomes, eGFRcrea, eGFRcysc, and UACR, we performed LD score regression using the *ldsc* software package (version 1.01)[36,37]. For this analysis, we used GWAS summary data from a subset of participants (N= 766,345 , excluding participants from 23andMe) from the GWAS on EA[23], and

GWAS summary from the European ancestry samples from the GWAS on kidney outcomes eGFRcrea, eGFRcysc, and UACR (see **Table 1**).

Table 1. Genome-wide association studies used for two-sample MR in the present study

Phenotype	Unit	N sample	Consortium	Reference
Years of schooling	SD (4.2 yrs)	1,131,881	SSGAC	*Lee et al. 2018*
eGFRcrea	*ln*-transformed	567,460	CKDGen	*Wuttke et al. 2019*
eGFRcysc	*ln*-transformed	24,061	CKDGen	*Gorski et al. 2017*
UACR	*int*	547,361	CKDGen	*Teumer et al. 2019*
UcreaC	µmol/L	361,194	UKBB	www.nealelab.is/uk-biobank/
SBP	mmHg	745,820	UKBB+ICBP	*Evangelou et al. 2018*

CKD, chronic kidney disease (defined as eGFRcrea < 60mL/min/1.73m^2); eGFR, estimated glomerular filtration rate; UACR, urinary albumin-to-creatinine ratio; UcreaC, urinary creatinine concentration; SBP, systolic blood pressure; SD, standard deviations; int, inverse normally transformed; SSGAC, Social Science Genetic Associations Consortium; CKDGen, Chronic Kidney Disease GENetics consortium; UKBB, UK Biobank; ICBP, International Consortium of Blood Pressure-genome wide association studies.

Systolic blood pressure

Finally, as a positive control to support our methods and findings, we performed MR analysis of the effect of gEA on blood pressure, a phenotype etiologically related to kidney function.[38] To this end, we leveraged GWAS summary data on systolic blood pressure (SBP)[39] and performed two-sample MR and 2SLS as described above.

RESULTS

Two-sample Mendelian randomization

After harmonization of data and removal of ambiguous palindromic SNPs, 1200 to 1210 of the 1271 SNPs remained for the MR analyses (**Table 2**). SNP details and their effects on EA and kidney outcomes are presented in **Supplementary Table S2**. MR scatterplots of SNP effects on EA and kidney outcomes are shown in **Figure 1**.

eGFRcrea and eGFRcysc

We found neither a relevant nor statistically significant effect of years of schooling on eGFRcrea; each SD higher gEA was associated with a 0.04% lower eGFRcrea (IVW MR estimate: B= -0.0004, 95%CI: -0.0037 to 0.0029, p= 0.805), which was non-significant. However, each SD higher gEA was associated with a higher eGFRcysc (3.2% increase, 95%CI: 1.9% to 4.6%, p=2.4x10^{-6}, **Table 2**). No heterogeneity in SNP effects was observed for eGFRcysc, which suggest low risk of bias due to horizontal pleiotropy (i.e. SNP affecting the outcome not only through the exposure, but

Table 2. Results from two-sample Mendelian randomization analysis

Outcome	Method	N SNPs	B	se	P-value	Cochran's Q (df)	Q P-value
*ln*eGFRcrea	IVW	1210	-0.0004	0.0017	0.8047	3042 (1209)	**1.71e-158**
	Weighted median	1210	0.0015	0.0017	0.3649		
	MR Egger	1210	0.0033	0.0059	0.5749		
	Simple mode	1210	-0.0012	0.0086	0.8870		
	Weighted mode	1210	0.0041	0.0068	0.5424		
	MR Egger intercept	-	-4.4×10^{-5}	6.8×10^{-5}	0.5108		
*ln*eGFRcysc	IVW	1203	0.0322	0.0068	**2.4×10^{-6}**	1217 (1202)	0.3742
	Weighted median	1203	0.0233	0.0102	**0.0224**		
	MR Egger	1203	0.0399	0.0243	0.1017		
	Simple mode	1203	0.0379	0.0497	0.4460		
	Weighted mode	1203	-0.0062	0.0467	0.8947		
	MR Egger intercept	-	-9.2×10^{-5}	0.0003	0.7416		
int UACR	IVW	1204	0.0596	0.0085	**2.47×10^{-12}**	2388 (1203)	**1.04e-80**
	Weighted median	1204	0.0435	0.0099	**1.09×10^{-5}**		
	MR Egger	1204	0.1143	0.0297	**1.25×10^{-4}**		
	Simple mode	1204	0.0091	0.0485	0.8506		
	Weighted mode	1204	0.0209	0.0401	0.6022		
	MR Egger intercept	-	-0.0007	0.0003	0.0547		
UcreaC (µmol/L)	IVW	1207	-0.1685	0.0103	**1.23×10^{-60}**	2673 (1206)	**1.09e-112**
	Weighted median	1207	-0.1432	0.0117	**2.38×10^{-34}**		
	MR Egger	1207	-0.1953	0.0357	**5.45×10^{-8}**		
	Simple mode	1207	-0.0666	0.0609	0.2743		
	Weighted mode	1207	-0.0727	0.0516	0.1587		
	MR Egger intercept	-	0.0003	0.0004	0.4340		
SBP (mmHg)	IVW	1184	-1.828	0.212	**5.55×10^{-18}**	6267 (1183)	**0**
	Weighted median	1184	-1.735	0.180	**3.49×10^{-23}**		
	MR Egger	1184	-1.528	0.760	**0.0445**		
	Simple mode	1184	-1.061	1.045	0.3080		
	Weighted mode	1184	-0.878	0.963	0.3684		
	MR Egger intercept	-	-0.004	0.009	0.6814		

Effects of EA on kidney outcomes per one standard deviation increase in years of schooling (~4.2 years). Abbreviations: CKD, chronic kidney disease; eGFRcrea, creatinine-estimated glomerular filtration rate; eGFRcysc, cystatin-C estimated GFR; UACR, urinary albumin-to-creatinine ratio; UcreaC, urinary creatinine concentration; SBP, systolic blood pressure; *ln*, natural log-transformed; *INT*, inverse normally transformed; B, effect estimate; se, standard errors; Q, heterogeneity statistic; df, degrees of freedom.

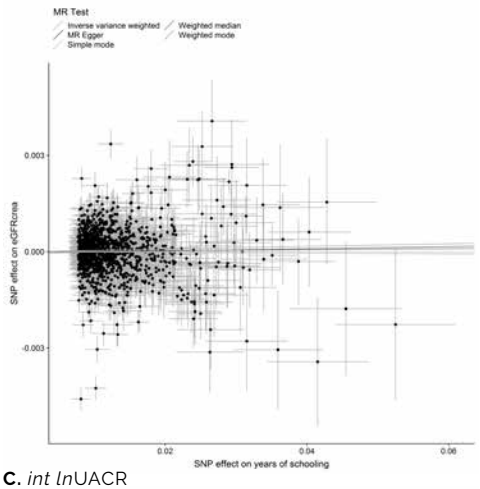
also through other traits, thereby violating MR assumptions and biasing the estimates). This was corroborated by sensitivity analyses that are robust to varying degrees of violations of MR assumptions (i.e. MR Egger, median and mode based methods, **Table 2**), as these yielded similar results. Additional sensitivity analysis was performed in a pruned dataset of 387 SNPs (**Supplementary Table S3**), for which we excluded SNPs based on strict criteria regarding inter-SNP correlation (linkage disequilibrium) and outlying effects (MR-PRESSO). This was done in an effort to minimize bias due to invalid SNPs, at the cost of statistical power. These also yielded largely similar effect estimates (**Supplementary Table S4**). SNPs for EA were more highly correlated with eGFRcysc than with EA, which resulted in non-significance of the MR Steiger test (r^2_{EA}=0.053, $r^2_{eGFRcysc}$=0.056, p=0.271, **Supplementary Table S5**). This suggested that EA was downstream of eGFRcysc. However, in a pruned dataset, the MR Steiger test was again significant (r^2_{EA}=0.020, $r^2_{eGFRcysc}$=0.016, p=0.027). Outlying SNP effects may indicate pleiotropy, and failing to account for these outliers may yield biased estimates. MR-PRESSO can be used to detect and account for any outlying SNPs. However, since there was no heterogeneity in SNP effects for eGFRcysc, MR-PRESSO did not detect any outliers for eGFRcysc (**Supplementary Table S4**).

UACR

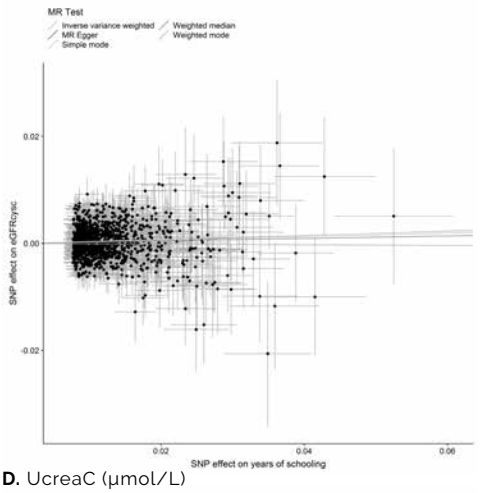
An effect of gEA on UACR was found, but this effect was not in the expected direction; each additional standard deviation in years of schooling (4.2 years) associated with an 0.06 increase in *int ln*UACR (IVW MR estimate: B=0.060, 95%CI: 0.043 to 0.077, p=2.5 $x10^{-12}$) (**Table 2, Figure 1**). Significant heterogeneity in SNP effect was detected (Cochran's Q=2388, df=1203, p=1.04 $x10^{-80}$), which suggests SNP pleiotropy. In case of unbalanced pleiotropy, effect estimates may be biased upwards or downwards. Egger intercept analysis can detect potential unbalanced pleiotropy. A suggestive bias towards a null effect was detected (Egger intercept: -6.54 $x10^{-4}$, 95%CI: -0.001 to 1.24 $x10^{-5}$, p=0.055). This means that unbalanced pleiotropy in SNP effects may have masked a stronger effect of EA on UACR. This is corroborated by a slightly larger effect-estimate in MR Egger sensitivity analysis that takes into account the imbalance (**Table 2**). Mode-based methods were directionally consistent but did not reach significance, possibly due to reduced power because of down-weighting of SNPs away from the mode. We observed no disproportionately influential SNPs in leave-one-out analyses (data not shown). The MR Steiger test[31] was significant (r^2_{EA}=0.052, r^2_{UACR}=0.005,

Figure 1. MR scatterplots

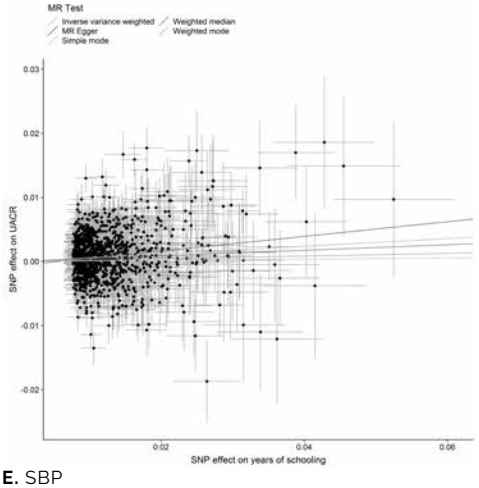
A. *lne*GFRcrea



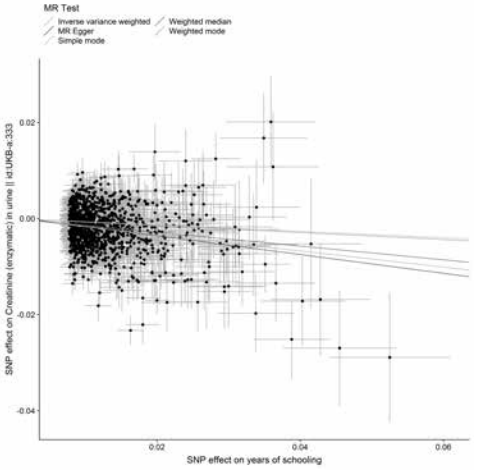
B. *lne*GFRcysc



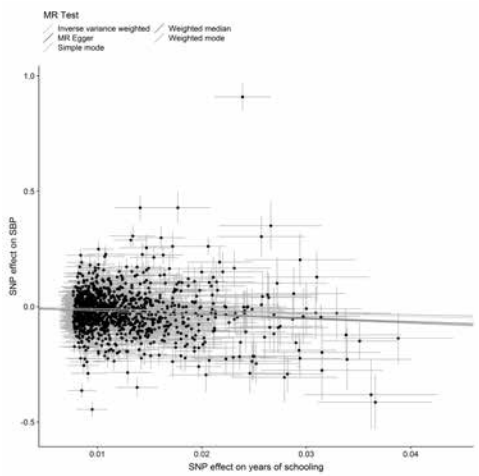
C. *int ln*UACR



D. UcreaC (µmol/L)

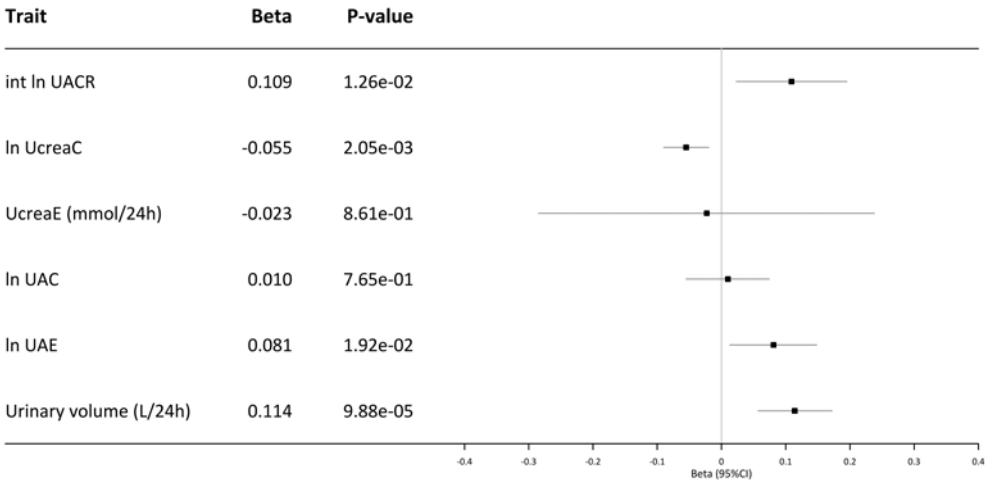


E. SBP



Mendelian randomization scatterplots. The X-axes represent SNP effects on standard deviations (4.2 years) of years of schooling The Y-axes represent SNP effects on **A.** natural log-transformed creatinine-estimated glomerular filtration rate (lneGFRcrea) **B.** natural log-transformed cystatin C-estimated glomerular filtration rate (lneGFRcysc); **C.** inverse normally transformed residuals of natural log-transformed urinary albumin-to-creatinine ratio (int lnUACR); **D.** urinary creatinine concentration (UcreaC) (µmol/L) and **E.** systolic blood pressure (SBP, mmHg)

Figure 2. Results of 2SLS analysis in Lifelines



Effects of 4.2 years of schooling on kidney traits obtained from 2-stage least squares (2SLS) regression in the Lifelines Cohort, adjusted for age, age², sex, and the first 10 genetic principal components. UACR, urinary albumin-to-creatinine ratio; UAC, urinary albumin concentration; UAE, urinary albumin excretion; UcreaC, urinary creatinine concentration; UcreaE, urinary creatinine excretion.

p<0.001) supporting that EA is causally upstream of UACR. Results from analyses with a more stringent SNP selection and additional MR-PRESSO outlier correction (**Supplementary Table S4**) yielded similar results.

Secondary analyses

Two-sample Mendelian randomization on urinary creatinine concentration

Given the unexpected direction of the effect of gEA on UACR, we decided to explore the effect of gEA on the creatinine component of UACR by performing a two-sample MR on UcreaC in the UK Biobank. Each 4.2-year higher gEA associated with 0.169 µmol/L lower UcreaC (95%CI: -0.189 to -0.148, p=1.23 x10^{-60}, **Table 2**). The significant MR Steiger test supported a causal direction from EA to UcreaC, given that SNPs for EA showed stronger correlation with EA than with UcreaC (r^2_{EA}=0.053, r^2_{UcreaC} =0.010, p<0.001). Given that UcreaC is the denominator in UACR, lower UcreaC results in higher UACR. These results suggest that bias due to the creatinine component of UACR could play a role. This hypothesis was further investigated in data from the Lifelines cohort that include 24hr urine samples that were not available in the UK Biobank, and only in few cohorts contributing to the CKDGen Consortium.

2-stage least squares analysis in Lifelines

Participant characteristics of the Lifelines Cohort study are presented in **Supplementary Table S6-S7**. The WGS for years of schooling was significantly correlated (r^2=0.04, p=4.5x10^{-108}) with EA in Lifelines, comparable to the correlation of the 1271 SNPs with years of schooling in two-sample MR (r^2 ≈ 0.05). Results of 2SLS analysis in Lifelines are presented in **Figure 2** and **Supplementary Table S8**. Similar to the two sample MR results, each 4.2-year increase in gEA was significantly causally associated with 0.109 higher *int ln*UACR (95%CI: 0.023 to 0.195, p=0.0126) and also confirmed by 0.081 higher 24h-*ln*UAE (95%CI: 0.013 to 0.148, p=0.0192). Each 4.2-year increase in gEA was associated with lower concentrations of Ucrea in 24h urine by 5.5% (2SLS B= -0.055, 95%CI: -0.090 to -0.020, p=2.05 x10^{-3}), consistent with the two-sample MR results from the UK Biobank UcreaC data. However, gEA did not affect 24hr urinary creatinine excretion (2SLS B= -0.023, 95%CI: -0.285 to 0.238, p=0.861). The finding that EA reduced UcreaC but not UcreaE points to dilution of urinary creatinine resulting from higher fluid intake in those with higher EA. This is supported by our observation that a 4.2-year increase in gEA indeed resulted in 0.114 L higher 24h urinary volumes (2SLS B= 0.114, 95%CI: 0.057 to 0.172, p=9.88 x10^{-5}).

Genetic correlations

To support the findings from our MR analysis, we computed genetic correlations between EA and kidney outcomes with LD score regression (see **Supplementary Table S9**). This method utilizes the complete GWAS data for both traits, whereas in MR only a genome-wide significant subset of SNPs associated with the exposure is used. Furthermore, this method is not sensitive to sample overlap[37] between the kidney outcomes GWAS data and the EA GWAS data, which may potentially have caused amplification of SNP effects and may therefore have resulted in biased MR estimates[40]. Genetic correlations of EA with eGFRcrea (r_g= -0.0129, p= 0.415), and eGFRcysc (r_g= 0.0925, p= 0.0144), and UACR (r_g= 0.1131, p= 2.09 x10^{-10}) were consistent with the MR estimates both in direction and in significance, further supporting our main findings.

SBP

As a positive control to support our methods and findings, we repeated each analysis on SBP. After harmonization, 1184 SNPs remained. In two-sample MR, an increase in gEA resulted in a decrease in SBP as expected (IVW MR estimate:

B= -1.83 mmHg per 4.2-year increase in gEA, 95%CI: -2.25 to -1.41, p= 5.55 x10^{-18}) (**Table 2**). Heterogeneity was detected, but no directional horizontal pleiotropy was observed (Egger intercept = -0.004, 95%CI: -0.020 to 0.013, p=0.681). The MR Steiger test was significant (r^2_{EA}=0.052, r^2_{SBP}=0.010, p<0.001), supporting an upstream role of EA. Sensitivity analyses robust to instrument pleiotropy yielded similar results, as did analyses with a more stringent SNP selection and MR-PRESSO outlier adjustment (**Supplementary Table S4**). This protective effect of gEA on SBP was corroborated in Lifelines, with a slightly larger effect size compared to the IVW MR estimate (2SLS B= -2.03 mmHg per 4.2-years increase in schooling, 95%CI: -3.28 to -0.775, p=1.51 x10^{-3}).

DISCUSSION

In this Mendelian randomization study, no convincing genetic support for a protective effect of EA on kidney outcomes was found. Although a protective effect of gEA was found on eGFRcysc, there were no effects of gEA on eGFRcrea, weakening the strength of the evidence for a protective effect of EA on kidney function. The effect of gEA on UACR was even in the opposite direction to what was expected. Taken together, these results challenge the notion that higher EA causally protects against CKD.

The absence of a protective effect of EA on CKD is surprising, given the large body of observational epidemiological evidence on this topic[5-7], and previous MR studies on the relation between EA and cardiovascular disease and CKD risk factors[13-16]. Contrary to expectations, we found that higher gEA resulted in higher UACR, a marker of kidney damage. Given that we observed significant effects of gEA on eGFRcysc and SBP, but not eGFRcrea, we hypothesized that a spurious association between EA and creatinine (e.g. through higher muscle mass in those with lower education) may have led to both the absence of effect on eGFRcrea and the unexpected detrimental effect on UACR. In a two-sample MR analysis we observed a relation between higher gEA and lower UcreaC, which could at least in part explain the observed higher UACR in those with higher gEA. To further explore this hypothesis, we performed a secondary analysis in individual participant data from the Lifelines Cohort, where we investigated effects of EA on both the creatinine and the albumin component of UACR in 24hr urine samples. We found that higher gEA was again significantly related to both higher UACR and lower UcreaC. However, when examining 24h excretions of creatinine as

outcome, no effect of gEA was observed, dismissing spurious associations with muscle mass as a source of confounding bias. The strong effect of higher gEA on higher 24h urinary volumes in Lifelines, points to higher fluid intake in those with higher EA. Higher fluid intake, however, is not a possible explanation of the counterintuitive effect of EA on UACR, as greater urinary volumes would lead to dilution of urinary creatinine as well as of urinary albumin. Consequently no effect on UACR is to be expected. Some have suggested that higher fluid intake may increase proteinuria[41,42] resulting in higher UACR, but this is controversial in light of accumulating evidence for a reno-protective effect of fluid intake[43,44].

We found genetic correlations of EA with eGFRcrea and eGFRcysc consistent with the present MR results. Furthermore, we found a modest, but highly significant, positive genetic correlation of EA with UACR, also consistent with our MR results. The positive genetic correlation means that genetic factors that correlate with higher EA also correlate with higher UACR, which is discrepant with the negative phenotypic correlation reported in literature[5]. As phenotypes depend both on additive genetic effects and environmental effects, the discrepancy between the genotypic and phenotypic correlation may possibly be explained by a strong environmental correlation between EA and UACR in the opposite direction, i.e. environmental factors likely exist that correlate both with higher EA and lower UACR. Further study is needed to identify responsible environmental and genetic factors and explain the counterintuitive deleterious effect of gEA on UACR.

This is the first MR study of the relation between EA and kidney outcomes. Strengths include highly precise SNP effect estimates from large GWASs on EA (1.1 million participants) and kidney outcomes (up to 567,140 participants), state-of-the art sensitivity analyses accounting for heterogeneity and pleiotropy in SNP effects, and 24h urine collections in Lifelines for our secondary analyses. Several limitations need to be addressed. First, MR IVW relies on untestable assumptions regarding pleiotropy of SNPs. We therefore applied a range of sensitivity analyses (i.e. MR Egger, median and mode-based methods, MR-PRESSO) that are robust to varying degrees of violation of these MR assumptions[27-30]. Essentially similar results were obtained. Second, eGFR is an approximation of kidney function based on serum creatinine as a marker. It is known that there can be marker induced bias, and that there is lower precision of GFR estimating equations in the higher ranges (>60 mL/min/1.73m^2)[45,46]. Therefore, effect estimates may be biased

towards the null. However, not only did we observe a lack of effect of gEA on eGFR, but also an opposite, deleterious effect on albuminuria, which strengthens our conclusion that there is no convincing genetic support for a protective effect of EA on CKD. Third, in previous meta-analysis of observational studies, the EA-CKD association was observed to be highly heterogeneous[5], possibly due to between-country differences in educational and health care systems. If the EA-CKD association is not consistent between study samples, this may result in bias towards the null in MR estimates. Fourth, SNP effect estimates were obtained from GWAS performed in populations of European descent from middle-to-high income countries. Therefore, generalizability to non-European ancestries and to low-income countries may not be possible. Lastly, sample overlap[40] (i.e. overlap in participants for different GWAS), assortative mating[47] (i.e. selective mating based on educational level), and genetic nurture effects[48] (i.e. indirect effects of non-transmitted parental alleles on offspring EA through rearing environment), may have caused amplification of SNP effects, which in turn would bias MR estimates away from the null. However, this is unlikely to have affected our main conclusions given that we observed no protective effect of EA on kidney outcomes.

The results of this study may have several implications. Our data suggest that, expected positive effects on general cardiovascular health notwithstanding, policies to optimize education may not reduce the burden of CKD in middle-to-high income communities of European descent. Moreover, our data also indicate that the consistent inverse association of EA with CKD that is found in epidemiological studies is possibly confounded, as we found no genetic support for a causal effect of EA on CKD. Further research is needed to ascertain which latent factors drive socioeconomic disparities in CKD. In that respect, it should be noted that future (genetic) epidemiological studies on kidney outcomes, including MR and GWAS, should consider possible marker (e.g. creatinine) induced bias in SNP effects.

The results of the present MR study indicate a null effect of EA on eGFR. Unexpectedly, genetic support for a counterintuitive, deleterious effect of EA on albuminuria was found, a finding that warrants further investigation. Taken together, we conclude that there is no convincing genetic support for a protective effect of EA on CKD.

ACKNOWLEDGEMENTS

The authors thank the CKDGen Consortium for providing full access to summary statistics for the eGFR and UACR phenotypes, and the ICBP consortium for providing access to summary statistics for the SBP phenotype. The authors wish to acknowledge the services of the Lifelines Cohort Study, the contributing research centres delivering data to Lifelines, and all the study participants. The Lifelines Biobank initiative has been made possible by funds from Fonds Economische Structuurversterking, Samenwerkingsverband Noord Nederland and REP (Ruimtelijk Economisch Programma).

CONFLICT OF INTEREST STATEMENT

The authors report no conflict of interest.

AUTHOR CONTRIBUTIONS

CHLT and *HS* designed the study; *CHLT* and *PJvdM* analyzed the data; *CHLT* drafted the work; *CHLT, PJvdM, UB, RTG, HS* contributed to the interpretation of data, as well as critical revision and final approval of the manuscript.

REFERENCES

1. Stevens PE, Levin A, Kidney Disease: Improving Global Outcomes Chronic Kidney Disease Guideline Development Work Group Members. 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):825-830.

2. Levey AS, Becker C, Inker LA. Glomerular filtration rate and albuminuria for detection and staging of acute and chronic kidney disease in adults: A systematic review. *JAMA*. 2015;313(8):837-846.

3. Chronic Kidney Disease Prognosis Consortium, Matsushita K, van der Velde M, et al. Association of estimated glomerular filtration rate and albuminuria with all-cause and cardiovascular mortality in general population cohorts: A collaborative meta-analysis. *The Lancet*. 2010;375(9731):2073-2081.

4. Gansevoort RT, Correa-Rotter R, Hemmelgarn BR, et al. Chronic kidney disease and cardiovascular risk: Epidemiology, mechanisms, and prevention. *The Lancet*. 2013;382(9889):339-352.

5. Vart P, Gansevoort RT, Joosten MM, Bultmann U, Reijneveld SA. Socioeconomic disparities in chronic kidney disease: A systematic review and meta-analysis. *Am J Prev Med*. 2015;48(5):580-592.

6. Zeng X, Liu J, Tao S, Hong HG, Li Y, Fu P. Associations between socioeconomic status and chronic kidney disease: A meta-analysis. *J Epidemiol Community Health*. 2018;72(4):270-279.

7. Weldegiorgis M, Smith M, Herrington WG, Bankhead C, Woodward M. Socioeconomic disadvantage and the risk of advanced chronic kidney disease: Results from a cohort study with 1.4 million participants. *Nephrol Dial Transplant*. 2019.

8. Vart P, Gansevoort RT, Crews DC, Reijneveld SA, Bultmann U. Mediators of the association between low socioeconomic status and chronic kidney disease in the united states. *Am J Epidemiol*. 2015;181(6):385-396.

9. Thio CH, Vart P, Kieneker LM, Snieder H, Gansevoort RT, Bültmann U. Educational level and risk of chronic kidney disease: Longitudinal data from the PREVEND study. *Nephrol Dial Transplant*. gfy361. 2018.

10. Davey Smith G, Ebrahim S. 'Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol*. 2003;32(1):1-22.

11. Davey Smith G, Hemani G. Mendelian randomization: Genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet*. 2014;23(R1):R89-R98.

12. Sekula P, Del Greco MF, Pattaro C, Kottgen A. Mendelian randomization as an approach to assess causality using observational data. *J Am Soc Nephrol*. 2016;27(11):3253-3265.

13. Tillmann T, Vaucher J, Okbay A, et al. Education and coronary heart disease: Mendelian randomisation study. *BMJ*. 2017;358:j3542.

14. Böckerman P, Viinikainen J, Pulkki-Råback L, et al. Does higher education protect against obesity? evidence using Mendelian randomization. *Prev Med*. 2017;101:195-198.

15. Gage SH, Bowden J, Davey Smith G, Munafo MR. Investigating causality in associations between education and smoking: A two-sample Mendelian randomization study. *Int J Epidemiol*. 2018;47(4):1131-1140.

16. Carter AR, Gill D, Davies NM, et al. Understanding the consequences of education inequality on cardiovascular disease: Mendelian randomisation study. *BMJ*. 2019;365:l1855.

17. Parikh NI, Hwang S, Larson MG, Meigs JB, Levy D, Fox CS. Cardiovascular disease risk factors in chronic kidney disease: Overall burden and rates of treatment and control. *Arch Intern Med*. 2006;166(17):1884-1891.

18. Burgess S, Scott RA, Timpson NJ, Smith GD, Thompson SG, EPIC-InterAct Consortium. Using published data in mendelian randomization: A blueprint for efficient identification of causal risk factors. *Eur J Epidemiol*. 2015;30(7):543-552.

19. Wuttke M, Li Y, Li M, et al. A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat Genet*. 2019;51(6):957-972.

20. Gorski M, van der Most PJ, Teumer A, et al. 1000 genomes-based meta-analysis identifies 10 novel loci for kidney function. *Scientific Reports*. 2017;7:45040.

21. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med*. 2009;150(9):604-612.

22. Inker LA, Schmid CH, Tighiouart H, et al. Estimating glomerular filtration rate from serum creatinine and cystatin C. *N Engl J Med*. 2012;367(1):20-29.

23. Lee JJ, Wedow R, Okbay A, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet*. 2018;50:1112-1121.

24. UNESCO Institute for Statistics. *International standard classification of education: ISCED 2011*. Montreal: UNESCO Institute for Statistics, 2012.

25. Hemani G, Zheng J, Elsworth B, et al. The MR-base platform supports systematic causal inference across the human phenome. *eLife*. 2018;7:e34408.

26. R Core Team. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria.

27. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol*. 2016;40(4):304-314.

28. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through egger regression. *Int J Epidemiol*. 2015;44(2):512-525.

29. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol*. 2017;46(6):1985-1998.

30. Verbanck M, Chen C, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet*. 2018;50(5):693.

31. Hemani G, Tilling K, Smith GD. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS genetics*. 2017;13(11):e1007081.

32. Scholtens S, Smidt N, Swertz MA, et al. Cohort profile: LifeLines, a three-generation cohort study and biobank. *Int J Epidemiol*. 2014;44(4), 1172-1180.

33. Klijs B, Scholtens S, Mandemakers JJ, Snieder H, Stolk RP, Smidt N. Representativeness of the LifeLines cohort study. *PloS one*. 2015;10(9):e0137203.

34. Nolte IM, van der Most, Peter J, Alizadeh BZ, et al. Missing heritability: Is the gap closing? an analysis of 32 complex traits in the lifelines cohort study. *Eur J Hum Genet*. 2017;25(7):877.

35. Thio CHL, Reznichenko A, van der Most PJ, et al. Genome-wide association scan of serum urea in european populations identifies two novel loci. *Am J Nephrol*. 2019;49(3):193-202.

36. Bulik-Sullivan BK, Loh P, Finucane HK, et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291.

37. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 2015.

38. National Kidney Foundation. K/DOQI clinical practice guidelines for chronic kidney disease: Evaluation, classification, and stratification. *Am J Kidney Dis*. 2002;39(2 Suppl 1):S1-266.

39. Evangelou E, Warren HR, Mosen-Ansorena D, et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat Genet*. 2018;50(10):1412.

40. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol*. 2016;40(7):597-608.

41. Clark WF, Kortas C, Suri RS, et al. Excessive fluid intake as a novel cause of proteinuria. *CMAJ*. 2008;178(2):173-175. doi: 10.1503/cmaj.070792.

42. Ravikovich E, Messersmith T, Mick G, McCormick K. Effect of oral fluid intake on urinary albumin excretion in diabetes mellitus. *J Diabetes Complications*. 2002;16(4):310-312.

43. Clark WF, Sontrop JM, Huang SH, Moist L, Bouby N, Bankir L. Hydration and chronic kidney disease progression: A critical review of the evidence. *Am J Nephrol*. 2016;43(4):281-292.

44. Guelinckx I, Vecchio M, Perrier ET, Lemetais G. Fluid intake and vasopressin: Connecting the dots. *Ann Nutr Metab*. 2016;68 Suppl 2:6-11.

45. Earley A, Miskulin D, Lamb EJ, Levey AS, Uhlig K. Estimating equations for glomerular filtration rate in the era of creatinine standardization: A systematic review. *Ann Intern Med*. 2012;156(11):785-795.

46. Levey AS, Inker LA, Coresh J. GFR estimation: From physiology to public health. *Am J Kidney Dis*. 2014;63(5):820-834.

47. Hartwig FP, Davies NM, Davey Smith G. Bias in mendelian randomization due to assortative mating. *Genet Epidemiol*. 2018;42(7):608-620.

48. Kong A, Thorleifsson G, Frigge ML, et al. The nature of nurture: Effects of parental genotypes. *Science*. 2018;359(6374):424-428. doi: 10.1126/science.aan6877.

Supplementary Material

CHAPTER 8

Box 1. List of abbreviations	
2SLS	two-stage least squares regression
CKD	chronic kidney disease
EA	educational attainment
eGFR	estimated glomerular filtration rate
gEA	genetically predicted educational attainment
GWAS	genome-wide association study
ISCED	international standard classification of education
IVW	inverse variance weighted
LD	linkage disequilibrium
MR	Mendelian randomization
MR-PRESSO	Mendelian randomization pleiotropy residual sum and outlier
SBP	systolic blood pressure
SNP	single nucleotide polymorphism
UAC	urinary albumin concentration
UACR	urinary albumin-to-creatinine ratio
UAE	urinary albumin excretion
UcreaC	urinary creatinine concentration
UcreaE	urinary creatinine excretion
WGS	weighted genetic score

Table S1. Educational attainment phenotype definition in the original genome-wide association study.

ISCED levels	**Definition**	**US years of schooling**
0	Pre-primary education	1
1	Primary education or first stage of basic education	7
2	Lower secondary or second stage of basic education	10
3	(Upper) secondary education	13
4	Post-secondary non-tertiary education	15
5	First stage of tertiary education (not leading directly to an advanced research qualification	19
6	Second stage of tertiary education (leading to an advanced research qualification, e.g. a Ph.D.	22

International Standard Classification of Education (ISCED) 1997 definitions with equivalent US years of schooling as defined in the original genome-wide association study (GWAS) on educational attainment.

Table S3. SNPs remaining after clumping procedure

rs10006235	rs1167827	rs1334297	rs2029401	rs34720381	rs57349798	rs72486027	rs7920624
rs10021733	rs11678980	rs13402497	rs2034631	rs34807077	rs57352738	rs7254263	rs7924036
rs10060023	rs11724690	rs13422673	rs2039204	rs35417702	rs575113	rs72622559	rs79265434
rs10080647	rs117398064	rs137079	rs2052285	rs35606437	rs5754753	rs72672052	rs7972246
rs1008078	rs118134876	rs1426619	rs2055940	rs35929923	rs57661533	rs72709560	rs79728014
rs10189857	rs11871429	rs1434630	rs2081652	rs363096	rs580652	rs72828517	rs7974852
rs10193498	rs11894424	rs143743568	rs2131167	rs3751331	rs590013	rs72829857	rs7977614
rs10215082	rs12054166	rs150537577	rs2179152	rs3809634	rs6043521	rs72944064	rs79994730
rs10460095	rs12076635	rs1505676	rs2183271	rs382196	rs6065080	rs72962169	rs8008382
rs10761251	rs12113634	rs1544	rs2199409	rs3859523	rs6065784	rs72993796	rs8024
rs10772644	rs12151248	rs1564347	rs2212430	rs3897821	rs61104616	rs730384	rs80257979
rs10773002	rs12375949	rs1603460	rs2216144	rs3948495	rs6123924	rs73055556	rs8030487
rs10773208	rs1245829	rs162445	rs2256965	rs401526	rs61527214	rs73191311	rs806816
rs10798888	rs12468040	rs1689510	rs2287838	rs401966	rs61798586	rs7321274	rs8097125
rs10799615	rs12477385	rs17048855	rs2290601	rs42210	rs61997667	rs7326331	rs891793
rs10805383	rs12503522	rs17110109	rs2297293	rs42302	rs62155350	rs73344830	rs912883
rs10844179	rs12506221	rs17144467	rs232496	rs4298514	rs62155873	rs73581580	rs9289300
rs10875121	rs12519073	rs17148998	rs2364544	rs4320563	rs62174974	rs736281	rs9359939
rs10931821	rs12524795	rs1717204	rs2365376	rs4328757	rs62179650	rs73648455	rs936496
rs10940921	rs12571549	rs17248751	rs2431023	rs4352658	rs62182994	rs737945	rs9373363
rs10951590	rs12591647	rs173003	rs2434672	rs4358081	rs62379838	rs743316	rs9375188
rs10979613	rs12602286	rs1738050	rs2436760	rs4369924	rs628993	rs74415461	rs9411331
rs10996167	rs12614263	rs17411339	rs2447097	rs4384309	rs6449503	rs7449561	rs9492774
rs11003463	rs12638072	rs17428076	rs2469226	rs4467547	rs6490618	rs7460106	rs9529146
rs11023764	rs12643771	rs1747714	rs2478208	rs4497562	rs6493265	rs746839	rs9556958
rs11030102	rs12646523	rs1747817	rs2496482	rs4673840	rs6557171	rs74747621	rs9616947
rs11076962	rs12670376	rs17489649	rs2517086	rs4675248	rs660001	rs74787922	rs9655780
rs11081529	rs12761761	rs17502934	rs252991	rs4719944	rs66568921	rs75033012	rs969512
rs11082011	rs12765185	rs175325	rs2545795	rs4726070	rs66721975	rs75177132	rs9853928
rs11121177	rs12789313	rs17551064	rs2554835	rs4757957	rs6690195	rs7575637	rs9859556
rs11130380	rs12875339	rs17563464	rs2570497	rs4766424	rs6697584	rs7603132	rs9882532
rs11138947	rs12888615	rs17565975	rs2725370	rs4778058	rs6704768	rs76076331	rs9886703
rs11157931	rs12957463	rs17598675	rs2764684	rs4787457	rs6731373	rs7617204	rs9927137
rs111852224	rs12967010	rs17604349	rs2838006	rs4793090	rs6812533	rs76235882	rs9927842
rs11211123	rs12981405	rs176218	rs28513882	rs4810894	rs68145588	rs76246107	rs9929556
rs11213482	rs13015496	rs17732878	rs2885198	rs4839155	rs6917154	rs7650602	rs9929762
rs112603734	rs13018640	rs178183	rs2923424	rs4846724	rs6924023	rs76577427	rs9933256
rs112806496	rs13050131	rs17882802	rs2929032	rs4848924	rs6969783	rs7672622	
rs1128956	rs13085461	rs182902112	rs2958182	rs488476	rs6977237	rs7683416	
rs113520408	rs13133213	rs1842713	rs2964199	rs4899012	rs6994287	rs76878669	
rs113615161	rs13145650	rs1861786	rs2964255	rs4904523	rs7012546	rs7692359	
rs114593137	rs13163845	rs1865955	rs2989476	rs4915735	rs7029718	rs77025239	
rs115000530	rs13197257	rs1866823	rs2998299	rs4977885	rs7040995	rs77609760	
rs11542663	rs1320139	rs1890132	rs3026996	rs4984541	rs7041702	rs77719387	
rs11588857	rs13212041	rs192436652	rs303752	rs55736314	rs7108020	rs7849487	
rs11598765	rs13240401	rs1933264	rs3111251	rs56099375	rs7127580	rs78648104	
rs11620355	rs13261773	rs1955250	rs34067381	rs56319902	rs7167688	rs7875078	
rs11657342	rs13266287	rs1991585	rs34098770	rs56391344	rs7171405	rs7894722	
rs11657979	rs1329125	rs2007655	rs34305371	rs56794817	rs717996	rs790647	
rs11663602	rs13327482	rs2011603	rs34394051	rs57204268	rs7226824	rs7910403	

List of 387 SNPs remaining after stringent LD clumping based on an r2-value of 0.001 according to the European samples of the 1000 Genomes project.

Table S4. Two-sample MR results after clumping procedure

exposure	outcome	method	nsnp	b	se	pval	Q	Q_df	Q_pval
years of schooling_clumped	*ln*eGFRcrea	MR Egger	372	0.0105	0.0099	0.2887			
years of schooling_clumped	*ln*eGFRcrea	Weighted median	372	0.0049	0.0028	0.0837			
years of schooling_clumped	*ln*eGFRcrea	Inverse variance weighted	372	0.0022	0.0027	0.4227	977.1	371	4.46E-56
years of schooling_clumped	*ln*eGFRcrea	Simple mode	372	0.0086	0.0089	0.3387			
years of schooling_clumped	*ln*eGFRcrea	Weighted mode	372	0.0075	0.0070	0.2814			
years of schooling_clumped	*ln*eGFRcrea	MR egger intercept	NA	-0.0001	0.0001	0.382			
years of schooling_clumped	*ln*eGFRcrea	MR-PRESSO outlier-corrected	361	0.0032	0.0024	0.1749			
years of schooling_clumped	*ln*eGFRcysc	MR Egger	368	0.0766	0.0397	0.0543			
years of schooling_clumped	*ln*eGFRcysc	Weighted median	368	0.0215	0.0163	0.1870			
years of schooling_clumped	*ln*eGFRcysc	Inverse variance weighted	368	0.0395	0.0109	0.0003	347.4	367	0.7612
years of schooling_clumped	*ln*eGFRcysc	Simple mode	368	-0.036	0.0568	0.5228			
years of schooling_clumped	*ln*eGFRcysc	Weighted mode	368	-0.030	0.0564	0.5948			
years of schooling_clumped	*ln*eGFRcysc	MR egger intercept	NA	-0.0005	0.0005	0.3314			
years of schooling_clumped	*ln*eGFRcysc	MR-PRESSO outlier-corrected	368	NA	NA	NA			
years of schooling_clumped	*int ln*UACR	MR Egger	370	0.1246	0.0499	0.0129			
years of schooling_clumped	*int ln*UACR	Weighted median	370	0.0519	0.0165	0.0016			
years of schooling_clumped	*int ln*UACR	Inverse variance weighted	370	0.0527	0.0139	0.0002	767.4	369	3.89E-30
years of schooling_clumped	*int ln*UACR	Simple mode	370	0.0494	0.0590	0.4030			
years of schooling_clumped	*int ln*UACR	Weighted mode	370	0.0494	0.0451	0.2735			
years of schooling_clumped	*int ln*UACR	MR egger intercept	NA	-0.001	0.0006	0.1342			
years of schooling_clumped	*int ln*UACR	MR-PRESSO outlier-corrected	364	0.0532	0.0131	5.99E-05			
years of schooling_clumped	UcreaC (µmol/L)	MR Egger	371	-0.194	0.0626	0.0021			
years of schooling_clumped	UcreaC (µmol/L)	Weighted median	371	-0.123	0.0195	2.47E-10			
years of schooling_clumped	UcreaC (µmol/L)	Inverse variance weighted	371	-0.153	0.0174	1.42E-18	923.4	370	4.03E-49
years of schooling_clumped	UcreaC (µmol/L)	Simple mode	371	-0.083	0.0722	0.253			
years of schooling_clumped	UcreaC (µmol/L)	Weighted mode	371	-0.088	0.0616	0.1521			
years of schooling_clumped	UcreaC (µmol/L)	MR egger intercept	NA	0.0005	0.0008	0.4990			
years of schooling_clumped	UcreaC (µmol/L)	MR-PRESSO outlier-corrected	362	-0.150	0.0161	9.93E-19			
years of schooling_clumped	SBP (mmHg)	MR Egger	363	-2.446	1.1838	0.0395			
years of schooling_clumped	SBP (mmHg)	Weighted median	363	-2.136	0.2763	1.08E-14			
years of schooling_clumped	SBP (mmHg)	Inverse variance weighted	363	-2.049	0.3219	1.94E-10	1734.3	362	1.16E-177
years of schooling_clumped	SBP (mmHg)	Simple mode	363	-3.945	1.1196	0.0005			
years of schooling_clumped	SBP (mmHg)	Weighted mode	363	-4.220	1.3493	0.0019			
years of schooling_clumped	SBP (mmHg)	MR egger intercept	NA	0.0051	0.0147	0.7279			
years of schooling_clumped	SBP (mmHg)	MR-PRESSO outlier-corrected	334	-2.229	0.255	1.17E-16			

Results of two-sample Mendelian randomization (MR) after stringent LD clumping based on a r^2-value of 0.001 according to the European samples of the 1000 Genomes project. *ln*eGFRcrea, *natural log-transformed* estimated glomerular filtration rate based on creatinine; *ln*eGFRcysc, *natural log-transformed* estimated glomerular filtration rate based on cystatin C; *int ln*UACR, *inverse normally transformed residuals of natural log-transformed* urinary albumin-to-creatinine ratio; UcreaC, urinary creatinine concentration; SBP, systolic blood pressure; MR-PRESSO, Mendelian randomization pleiotropy residual sum and outlier.

Table S5. MR Steiger test for causal direction

Exposure	N SNPs	Outcome	SNP $r^2_{exposure}$	SNP $r^2_{outcome}$	inferred causal direction	Steiger Pvalue
EA	1210	eGFRcrea	0.0527	0.0060	exposure causes outcome	0.0000
EA clumped	372	eGFRcrea	0.0202	0.0019	exposure causes outcome	0.0000
EA	1203	eGFRcysc	0.0528	0.0560	outcome causes exposure	0.2711
EA clumped	368	eGFRcysc	0.0203	0.0164	exposure causes outcome	0.0273
EA	1204	UACR	0.0524	0.0048	exposure causes outcome	0.0000
EA clumped	370	UACR	0.0202	0.0015	exposure causes outcome	0.0000
EA	1207	UcreaC	0.0526	0.0103	exposure causes outcome	0.0000
EA clumped	371	UcreaC	0.0203	0.0035	exposure causes outcome	0.0000
EA	1184	SBP	0.0516	0.0097	exposure causes outcome	0.0000
EA clumped	363	SBP	0.0199	0.0028	exposure causes outcome	0.0000

MR Steiger test for causal direction. The exposure is inferred to be causally upstream of the outcome in case the instrumental SNPs are more highly correlated to the exposure than to the outcome. SNP, single nucleotide polymorphism; EA, educational attainment; eGFRcrea, estimated glomerular filtration rate based on creatinine; eGFRcysc, estimated glomerular filtration rate based on cystatin C; UACR, urinary albumin-to-creatinine ratio; UcreaC, urinary creatinine concentration; SBP, systolic blood pressure.

Table S6. Lifelines Cohort study participant characteristics

	Proportion	Mean	SD	25th%	Median	75th%	missings
N		12675					
Age (years)		48.95	10.70	41	48	55	0
Females	58%						0
Years of schooling		13.50	4.18	10	13	20	0
WGS years of schooling		16.01	0.27	15.83	16.01	16.19	0
eGFRcrea (mL/min/1.73m^2)		92.53	14.39	82.74	93.65	103.50	7
UACR (mg/mmol)		0.94	7.99	0.16	0.30	0.56	58
24h UAC (mg/L)		6.60	49.60	1.1	2	4	58
24h UAE (mg/24h)		11.92	104.33	2.07	3.61	6.84	84
24h UcreaC (mmol/L)		7.83	3.82	5	6.9	9.8	58
24h UcreaE (mmol/24h)		13.10	4.38	10.08	12.42	15.72	83
urinary volume (mL/24h)		1.89	0.67	1.401	1.828	2.322	81
SBP (mmHg)		128.49	15.61	118	127	138	23

WGS, weighted genetic score; eGFRcrea, estimated glomerular filtration rate based on creatinine; eGFRcysc, estimated glomerular filtration rate based on cystatin C; UACR, urinary albumin-to-creatinine ratio; UAC, urinary albumin concentration; UAE, urinary albumin excretion; UcreaC, urinary creatinine concentration; UcreaE, urinary creatinine excretion; SBP, systolic blood pressure.

Table S7. Educational attainment phenotype in the Lifelines Cohort study				
What is your highest completed education?	**Level**	**US years of schooling**	**N**	**%**
1) No Education (not finished elementary school)	ISCED 0	1	77	1
2) Lower education (elementary school)	ISCED 1	7	411	3
3) Lower or preparatory applied education (e.g. lower technical school, lower vocational education in business and administration , preparatory middle-level applied education)	ISCED 2	10	2273	18
4) Middle general continued education(e.g. further extended primary education, (further) extended primary education ,middle-level applied education-short , preparatory middle-level applied education theoretical)	ISCED 2	10	2079	16
5) Middle-level applied education(e.g. middle-level applied education-long, middle level applied/technical training, upper vocational education in business and administration)	ISCED 3	13	3546	28
6) Higher general and preparatory education(e.g. higher general continued education, preparatory scientific education, higher commoner's school)	ISCED 3	13	1026	8
7) Higher professional education or pre university education(e.g. higher professional education, higher level applied/technical training, higher vocational education in business and administration)	ISCED 5	20˙	2683	21
8) Scientific education (university)	ISCED 5	20˙	580	5
Lifelines educational attainment questionnaire, mapped to the International Standard Classification of Education (ISCED) 1997, with equivalent US years of schooling. ˙20 years imputed instead of 19 given that the questionnaire does not distinguish between ISCED level 5 and 6.				

Table S8. 2-stage least squares analysis in the Lifelines Cohort study							
Outcome	**B**	**se**	**95%CI LL**	**95%CI UL**	**t**	**Pvalue**	**N**
*int ln*UACR (LOD imputed)	0.109	0.044	0.023	0.195	2.494	1.26E-02	12617
ln UAC (LOD imputed)	0.010	0.033	-0.055	0.074	0.299	7.65E-01	12617
ln UAE (LOD imputed)	0.081	0.034	0.013	0.148	2.343	1.92E-02	12552
ln UcreaC	-0.055	0.018	-0.090	-0.020	-3.083	2.05E-03	12617
UcreaE (mmol/24h)	-0.023	0.133	-0.285	0.238	-0.176	8.61E-01	12552
Urinary volume (L/24h)	0.114	0.029	0.057	0.172	3.895	9.88E-05	12594
SBP (mmHg)	-2.027	0.639	-3.279	-0.775	-3.173	1.51E-03	12652
Results of 2SLS analysis in Lifelines. Estimates are effects of a 4.2 years increase in years of schooling on residuals of outcomes adjusted for age, age^2, age^3, sex, and the first 10 genetic principal components. All urinary markers were determined in 24h urine collections. UACR, urinary albumin-to-creatinine ratio; UAC, urinary albumin concentration; UAE, urinary albumin excretion; UcreaC, urinary creatinine concentration; UcreaE, urinary creatinine excretion; SBP, systolic blood pressure; LOD, (lower) limit of detection; *ln*, natural log-transformed; *int*, inverse normally transformed.							

Table S9. Genetic correlations

trait 1	trait 2	r_g	se	z	P	h2_obs	h2_obs_se	h2_int	h2_int_se	gcov_int	gcov_int_se
Years of schooling	eGFRcrea	-0.0129	0.0158	-0.8151	0.415	0.0754	0.0055	1.0044	0.0233	0.0072	0.0077
Years of schooling	eGFRcysc	0.0925	0.0378	2.4468	0.0144	0.1675	0.0667	0.9523	0.014	0.0175	0.0062
Years of schooling	UACR	0.1131	0.0178	6.3547	2.09E-10	0.0434	0.0021	0.9533	0.0098	2.10E-06	0.0069
eGFRcrea	eGFRcysc	0.5354	0.157	3.4095	0.0007	0.1665	0.0669	0.9532	0.0139	0.0943	0.0095
eGFRcrea	UACR	0.3359	0.0279	12.0265	2.58E-33	0.0434	0.0021	0.9527	0.0088	0.0147	0.0078
eGFRcysc	UACR	0.2091	0.0615	3.4009	0.0007	0.0423	0.0023	0.9687	0.0118	-0.0096	0.0059

Genetic correlations (r_g) between traits were calculated using linkage disequilibrium (LD)-score regression implemented in the *ldsc* software package (version 1.01). GWAS summary statistics of European ancestry (sub) samples were used. To minimize bias due to poor imputation, summary statistics were restricted to HapMap3 SNPs. Single nucleotide polymorphisms (SNPs) were then filtered for missing values, minor allele frequency ≤ 0.01, ambiguous SNPs, duplicate SNPs, and SNPs in the major histocompatibility (MHC) region on chromosome 6. Pre-computed LD-scores for Europeans were used (available online at https://data.broadinstitute.org/alkesgroup/LDSCORE/eur_w_ld_chr.tar.bz2). GWAS summary statistics of years of schooling lacked a sample size column; a sample size of N=766,345 was therefore assumed for each SNP.

General discussion

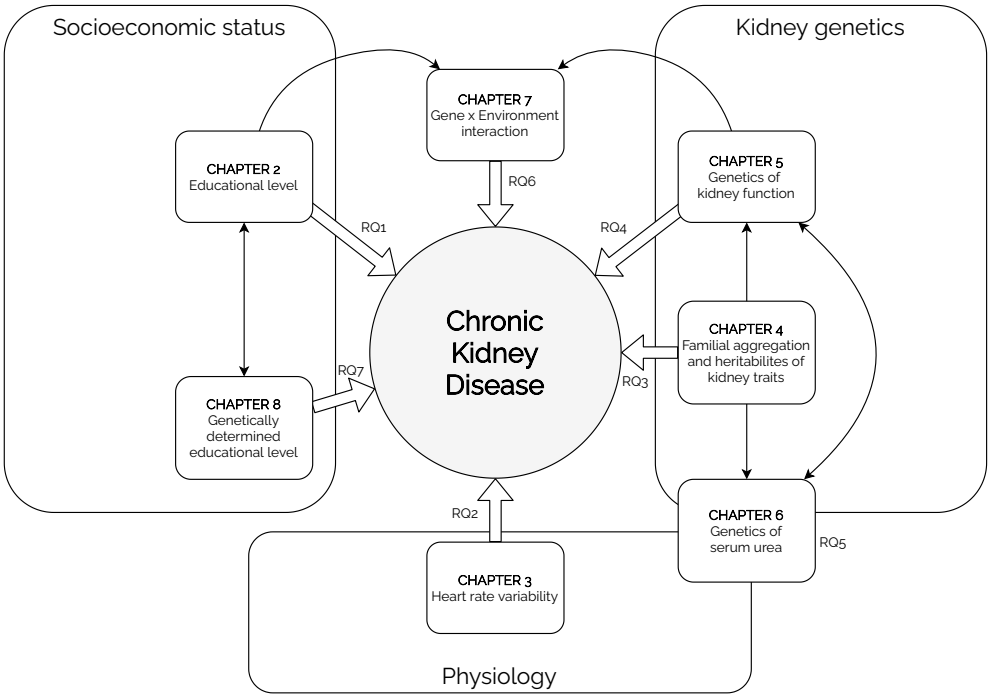
CHAPTER 9

GENERAL DISCUSSION

In this thesis, I examined mechanisms that influence risk of chronic kidney disease (CKD). Of particular interest were socioeconomic disparities in CKD, and how knowledge of the genetics underlying CKD can help in understanding these disparities. In a range of studies, I applied a variety of traditional epidemiological methods as well as genetic epidemiological designs and concepts, such as genetic risk scores (GRS), a family design, a genome wide association study (GWAS), a gene x environment interaction study, and Mendelian randomization (MR). In **Figure 1**, I provide a graphical overview of the different chapters, the research questions (RQs 1-7), and their interrelationship within this thesis.

This chapter contains the general discussion of my findings. First, I reflect on these findings in a broader context, and comment on the methods applied in this thesis. Then, I discuss future perspectives and the implications of my findings for public health, as well as clinical and research practice.

Figure 1. Graphic representation of the research questions (RQs) and their interrelationship in the context of this thesis. White arrows represent hypothesized direction of effect. Black arrows reflect overlap between chapters in use of data, methods, and/or results.



PART I: EVALUATING THE EFFECT OF SOCIOECONOMIC STATUS AND AUTONOMIC DYSFUNCTION ON RISK OF CKD

Chapter 2: Educational attainment is associated with risk of chronic kidney disease in the general population1

The predominance of studies on the relation between socioeconomic status) and CKD are based on cross-sectional, US data[2,3]. In **Chapter 2**, I present longitudinal data from the PREVEND cohort study, a Dutch community-based observational study with serial follow-up, where I used educational attainment (EA) as an indicator of socioeconomic status. In this sample, participants with low EA were at a 25% higher risk of developing CKD, and on average had a 0.15 mL/min/1.73m^2 faster eGFR decline per year compared to those with high EA, taking into account age and sex. Further analysis suggested a mediating role for modifiable factors such as hypertension, diabetes, and anthropometric indices, corroborating several cross-sectional findings from a previous mediation analysis[4].

An important finding from this study is the suggested mediating role of poor diet, i.e. low EA is associated with CKD through poor diet, in particular low potassium intake (i.e. few fruits or vegetables[5]). A role of poor diet in exacerbating CKD risk has long been proposed[6]. However, there is a paucity of data as only one previous study formally tested mediation by poor diet[4], but that study was limited because only questionnaire data on fruit intake was available. In contrast, I assessed diet by examining 24h urine excretions of nutrients as objective measures of nutrient intake (i.e. sodium, potassium, magnesium, and protein).

This study adds to the literature by providing insights into the mechanisms underlying the EA-CKD relation. In my models, the addition of potential mediators did not completely explain the association between EA and eGFR decline. This suggests the existence of non-traditional intermediate factors in the EA-CKD association.

Box 1. Abbreviations
CKD = chronic kidney disease
eGFR = estimated glomerular filtration rate
EA = educational attainment
GRS = genetic risk score
GWAS = genome wide association study
MR = Mendelian randomization
SNP = single nucleotide polymorphism

Future study may focus on identifying these non-traditional factors. In addition, further study of the intermediate pathways is necessary. Importantly, establishing the interrelationship between the different mediating factors (e.g. the effect of poor diet on body-mass index and hypertension, and body-mass index on diabetes, etc.) may help in prioritizing targets for public health interventions to mitigate socioeconomic disparities in CKD.

Chapter 3: Low heart rate variability does not precede chronic kidney disease in the general population[7]

Given that non-traditional risk factors may play a role in the development of CKD, I examined the effect of low heart rate variability, as an index of autonomic dysfunction, on CKD incidence (**Chapter 3**). Low heart rate variability occurs in the presence of an imbalance in autonomic function, when parasympathetic function is reduced relative to sympathetic function. Potentially, autonomic dysfunction leads to renal damage through changes in renal hemodynamics, with some evidence for such a detrimental effect in animal models. This effect of low heart rate variability on kidney health may exist in humans as well. Previously, a community based study in the US reported associations of low heart rate variability with CKD related hospitalization and end-stage renal disease[8]. Thus, the expectation was that low heart rate variability precedes new-onset CKD. In longitudinal analyses of data of 4605 subjects participating in the PREVEND cohort study, I observed a 50-100% higher risk of incident CKD for participants in the lowest quartile of heart rate variability measures, relative to those in the upper three quartiles. However, this association appeared to be completely driven by higher age in those with low heart rate variability and CKD. Thus, I could not corroborate a relation between low heart rate variability and CKD incidence. Rather, in post-hoc analyses, I found evidence suggesting an effect in the opposite direction (i.e. reverse causation), that is, CKD resulting in low heart rate variability, given that low heart rate variability was associated with kidney function only in those with CKD.

PART II: GENETICS OF KIDNEY FUNCTION AND THE TRANSLATION TO CLINICAL AND RESEARCH PRACTICE

Chapter 4: The heritability of kidney traits is considerable, and family history is an important determinant of CKD in the general population

Using the unique multi-generational family design of Lifelines, I estimated the heritability, i.e. the contribution of genetic factors to inter-individual variation in

a number of kidney traits. I observed considerable heritability of eGFRcrea (44%), urinary albumin excretion (20%), and serum urea (31%), among others. Furthermore, I computed the relative risk of developing CKD conditional on affected relatives. Here, I found that compared to the general population, the risk of having CKD for an individual is three times higher in case of a first-degree relative with CKD.
This study is the largest study of familial aggregation of kidney traits to date. An important observation in this study is that a positive family history strongly increases risk of CKD, suggesting a genetic component to kidney health. Furthermore, the heritability estimates provide an upper bound to the proportion of variance in kidney traits that can be explained by genetic factors. Future studies may focus on identifying these genetic factors.

Chapter 5: A genetic risk score based on 53 SNPs associated with eGFRcrea is a useful genetic proxy of kidney function, but possibly not of CKD susceptibility[9]
In addition to traditional clinical risk factors and lifestyle factors, genetic factors play a role in CKD. Recent genome-wide association studies (GWAS) have identified genetic variants associated with the CKD defining traits, eGFR[10] and albuminuria[11]. To date, GWAS for eGFR have been the most successful in terms of number of discovered variants. In the most comprehensive GWAS at the time, 53 SNPs were reported to have associations with eGFR estimated from serum creatinine (eGFRcrea)[12]. Each of these SNP effects were small and therefore unlikely to have meaningful clinical impact. However, it is possible to aggregate all SNP effects into one composite genetic risk score (GRS). Such a GRS may have utility in clinical practice as a risk stratification tool, and in research as a proxy for genetic predisposition. In **Chapter 5**, I evaluated a GRS based on these 53 eGFRcrea SNPs. Using data from 3649 subjects from the PREVEND cohort study, I found modest but robust associations of the GRS with eGFRcrea outcomes. These results were validated using eGFR estimated from cystatin C (eGFRcysc) rather than creatinine; similar associations of the GRS with eGFRcysc were found. This is important, given that eGFRcrea may in part reflect muscle mass rather than kidney function per se. Another important finding is that the GRS was not associated with albuminuria, and had an effect on eGFR independent of the renal risk factors, body-mass index, smoking, hypertension, diabetes, high cholesterol, and history of cardiovascular disease. This indicates that the GRS is a true genetic proxy of kidney function, not of kidney damage or kidney risk factors. However, the GRS only explained 1% in eGFR variance in PREVEND. Furthermore, longitudinal data were inconclusive: the

GRS was not significantly associated with eGFR decline, nor with incidence of CKD conditional on baseline eGFR. I therefore conclude that the GRS is unlikely to have a meaningful role in risk prediction of CKD. However, as a true genetic proxy of kidney function, the GRS may have utility in population level research, and in designs such as Mendelian randomization.

Chapter 6: Genome-wide association study of serum urea in Europeans identifies two novel genetic loci[13]

In **Chapter 5**, I used two different biomarkers for eGFR, namely serum creatinine and serum cystatin C. In **Chapter 6**, I investigated the genetics of serum urea (also known as blood urea nitrogen, BUN, when only the nitrogen component of urea is assayed). Serum urea is another commonly used, diagnostic marker for kidney function that was shown to be heritable in **Chapter 4**. Genetic data on this trait has been derived predominantly from East-Asian populations, where SNPs at 13 loci were known at the time[14-17]. Only few studies investigated this trait in European populations. These studies were either unsuccessful in finding associations[18], or lacked a replication phase for the five associations that were found. I therefore performed the first meta-analysis of GWASs on serum urea in European populations, with a gene discovery phase in 13,312 participants from the Lifelines Cohort, and built-in replication of the findings in 7379 participants from three community based cohorts (PREVEND, NESDA, and EGCUT). I identified replicable associations of six SNPs at the genome-wide level ($p < 5 \times 10^{-8}$), of which two were novel findings (rs2003313 on chromosome 11 near *POU2AF1*, and rs998394 on chromosome 3 in *ADAMTS9-AS2*).Furthermore, all SNPs previously identified in either East-Asians or Europeans were replicated, except for SNPs at three loci that are potentially specific to East-Asians.

I then aimed to identify potential causal genes involved in the pathways underlying urea metabolism and explore potential relevance to kidney function. Of the six identified SNPs, two were novel. Bioinformatics analysis of these two novel loci did not yield a clear relation to urea metabolism or kidney function, and thus, additional functional work is needed. An interesting candidate locus with regards to kidney function and disease is the *MUC1* locus. In kidney biopsy specimens, I found one of the identified SNPs, rs914615, to be an expression quantitative trait locus (eQTL) for *MUC1*, i.e. SNP rs914615 is linked to *MUC1* gene expression. Other SNPs tagging the *MUC1* locus have been consistently associated with serum urea in

previous studies[16]. It is a locus with potential clinical relevance for several reasons: it is involved in ion channels and electrolyte balance; aberrant activation of *MUC1* has been related to CKD development and; a frameshift mutation in *MUC1* causes medullar cystic kidney disease type 1[19,20]. Furthermore, a recent GWAS reported associations of albuminuria with SNPs that tag the *MUC1* locus[11]. Finally, in a recent study, differential expression of *MUC1* in the kidney was suggested to affect eGFR[21], adding evidence for a role of this locus in the development of kidney disease.

Next, I investigated the overlap of my findings with genetic data on kidney function. Overlap can be expected between serum urea and creatinine-based indices of kidney function, as serum levels of both urea and creatinine are influenced by kidney function. In a previous family analysis, a genetic correlation (r_g=0.56) was found between urea and creatinine[18], suggesting pleiotropy between these two traits. The positive direction of the genetic correlation indicates that shared genetic factors between urea and creatinine affect serum levels of both in the same direction (i.e. higher urea is genetically correlated with higher creatinine). Adding to this evidence is my finding that the 53 SNPs associated with eGFRcrea were enriched for associations with serum urea; 14 out of 53 eGFRcrea SNPs were also associated with serum urea, much more than could be expected based on random chance. Furthermore, a GRS based on these 53 eGFRcrea SNPs (the same GRS as in **Chapter 5**) was modestly but significantly associated with serum urea. The effect of this GRS was attenuated after adjustment for eGFRcrea, suggesting that the GRS indeed affects serum urea levels through kidney function.

Notwithstanding these statistically significant results, the clinical utility of these GWAS data on serum urea is rather limited. Together, the identified genetic variants explained no more than 0.56% of serum urea variation. My findings do, however, generate hypotheses for two novel loci (*POU2AF1* and *ADAMTS9-AS2*) with regards to urea and kidney function biology that may be investigated in functional research. Furthermore, the GWAS results on serum urea may be utilized in validating proposed kidney function loci: if a genetic variant is truly a marker of kidney function, the variant is expected to be related to both higher eGFRcrea and lower serum urea (or vice versa). This is exemplified in the most recent GWAS on eGFRcrea[10], in which the authors used GWAS results on BUN (the nitrogen component of urea) as a positive control to validate their findings.

PART III: UTILIZING GENETICS TO EXPLAIN SOCIOECONOMIC

DISPARITIES IN CHRONIC KIDNEY DISEASE

Chapter 7: Low educational attainment amplifies genetic risk of CKD in the general population

In **Chapters 7** and **8**, I applied the knowledge gained in previous chapters to integrate genetic methods with traditional social epidemiological methods. In **Chapter 7**, using data from the PREVEND cohort study, I present evidence for an amplifying effect of low EA on genetic risk of low eGFR. This finding was most pronounced in longitudinal analysis, where I observed an interaction between low EA and a high GRS. This interaction resulted in a more rapid rate of eGFR decline for those with both a high GRS and a low EA with a departure from additivity, meaning that the joint effects of a GRS and EA are larger than the sum of their main effects. Furthermore, these results suggest that high EA the genetic risk of eGFR decline, given that no apparent effect of a GRS was found in this group. This interaction could not entirely be explained by traditional risk factors (body-mass index, smoking, cholesterol, blood pressure, and glucose), suggesting the existence of unmeasured mediating factors whose influence is not captured by traditional factors.

These results add to the literature, as these are the first to provide evidence of a gene-environment interaction effect on kidney outcomes resulting from a modifying effect of EA. Importantly, I found that genetic risk of CKD is equally distributed across strata of EA, suggesting that there is no selection on kidney risk variants in those with low EA. Hence, the higher risk of CKD in those with low EA is attributable to an amplified effect of a GRS due to low EA itself or due to downstream effects of low EA. The results plead against genetic determinism in CKD, i.e. the risk of developing disease is not predetermined based on one's genes. Given that the interaction effect was rather modest and only accounted for ~0.1% of explained variance in rate of eGFR decline, its utility in risk stratification of individuals is negligible. However, if the effect is proven to be replicable in other samples, some benefit is to be expected from population level intervention on EA and its modifiable downstream effects in mitigating genetic risk of eGFR decline. Furthermore, although this study was sufficient powered to identify interaction effects on continuous outcomes, larger numbers are needed to assess whether the interaction effect results in increased risk of CKD, based on clinical cut-off values. Finally, the results warrant further characterization of the

mediating pathways between EA and CKD, and the specific genes involved in these pathways.

Chapter 8: The association between educational attainment and CKD may be confounded

The results in **Chapter 2** suggest a reno-protective effect of higher EA, as higher EA was associated with slower eGFR decline and lower CKD incidence, However, it is uncertain whether this association represents a true causal relation due to the observational nature of the data. In **Chapter 8**, I applied a Mendelian randomization method that uses genetic proxies for EA to minimize bias, thereby strengthening causal inference. For the two-sample MR analysis, I obtained data on 1271 SNPs with known effects on years of schooling[23], and interrogated the effect of these SNPs in genetic summary data from the CKDGen Consortium on eGFRcysc, eGFRcrea, and albuminuria (urinary albumin-to-creatinine ratio). I found that each one sd (4.2 years) higher EA was associated with a 3.2% higher eGFRcysc, consistent with my prior hypothesis of a protective effect. However, I found a null effect on eGFRcrea. A higher EA was even associated with higher urinary albumin-to-creatinine ratio, suggesting that higher EA results in kidney damage. To further investigate this counterintuitive finding, I performed secondary analyses in individual-level data of the Lifelines cohort, in which more detailed albuminuria data are available. I computed a genetic score based on the 1271 SNPs for years of schooling, and used this score as a genetic proxy for years of schooling. The counterintuitive detrimental effect of EA on urinary albumin-to-creatinine ratio found in the two-sample MR analysis was also observed using data of the Lifelines cohort. Here, I corroborated that this was due to higher urinary albumin excretion and not due to lower urinary creatinine excretion, thus not an artifact of lower muscle mass. This suggests that higher EA indeed leads to increased albuminuria.

Given the existing evidence on the protective effects of EA on cardiovascular health[24,25], protective effects on renal health were expected. However, I found inconsistent effects of EA on eGFRcrea and eGFRcysc, and an unexpected detrimental effect on urinary albumin-to-creatinine ratio and urinary albumin excretion. Future study may investigate what mechanisms explain this apparent detrimental effect on albuminuria. Based on these results, I conclude that there is insufficient genetic evidence for a protective causal effect of EA on kidney health. Thus, future studies on disparities in CKD may investigate other potentially

causative socioeconomic factors such as income, occupation and occupational exposures, social deprivation, or area-level indicators of socioeconomic status.

METHODOLOGICAL CONSIDERATIONS

General comments

Important strengths of this thesis include its use of multiple datasets, and the multidisciplinary approach to analyzing these data. I combined the expertise from the fields of nephrology, social epidemiology, and genetic epidemiology. This combination resulted in a wide range of analytic approaches: traditional epidemiological methods in **Chapters 2** and **3**, a family study in **Chapter 4**, genetic risk score application in **Chapters 5** through **8**, a GWAS in **Chapter 5**, a gene-environment interaction study in **Chapter 7**, and a Mendelian randomization study in **Chapter 8**. In this section, I describe the most important data sources and comment on the methods applied in this thesis.

Data sources

The research questions in this thesis were addressed using data from a number of existing sources. Here, I discuss the data sources that contributed most to this thesis, namely the Prevention of REnal and Vascular ENdstage Disease (**PREVEND**) cohort study, the Lifelines Cohort study and Biobank (**Lifelines**), and the Chronic Kidney Disease Genetics (**CKDGen**) consortium.

Data from the **PREVEND** cohort study[26] was used for **Chapters 2**, **3**, **5**, **6**, and **7**, while it contributed in part to **Chapter 8**. This prospective, observational cohort was sampled from the general population of the city of Groningen, the Netherlands. It was originally initiated to study the natural course of albuminuria and its association with renal and cardiovascular outcomes. PREVEND is ideally suited for investigating kidney outcomes due to its substantial follow-up duration (five consecutive examination rounds between 1997 and 2010). Importantly, PREVEND allows for precise measurement of kidney function and damage, with serum creatinine, serum cystatin C, and urinary albumin excretion being available. Furthermore, two 24h urine collections per examination round were available, allowing for optimal evaluation of albuminuria. The baseline sample consisted of ~8600 participants, of which a random sample of ~3500 was genotyped with a genome-wide array.

For **Chapters 4, 6** and **8**, data from **Lifelines**[27] was used. Lifelines is a large,

population-based prospective cohort study sampled from the Netherlands' three northernmost provinces (Groningen, Friesland, and Drenthe). From 2006 to 2013, ~165,000 participants were included and extensively phenotyped. Currently, genotype data is available for ~13,500 participants, which were included for analysis in **Chapters 6** and **8**. For the baseline measurement, 24h urine collections were available, which allows for exact evaluation of urinary creatinine and urinary albumin excretion. However, only two surveys were currently available; follow-up data is therefore limited. Another limitation of Lifelines with regards to kidney research are that measurements of serum creatinine are available but not of serum cystatin C, and that there are no follow-up data on urinary albumin excretion, a determination of urinary albumin was discontinued after the first 60,000 participants were measured at the baseline assessment. For **Chapter 4**, I exploited the multi-generational design in Lifelines to perform the largest family study on kidney outcomes to date, with >29,000 families and up to 4 generations per family.

Another important data source was the Chronic Kidney Disease Genetics consortium (**CKDGen**). CKDGen is an international collaborative effort to investigate the genetics of kidney outcomes. For **Chapter 5**, **6**, and **7**, I constructed a genetic risk score based on the then-known 53 or 63 genome-wide significant SNPs reported by CKDGen in 2016 and 2017, respectively[12,28]. In **Chapter 8**, I used summary statistics derived from a more recent and comprehensive GWAS meta-analysis on $eGFR_{crea}$[10] as well as the latest GWAS meta-analysis on urinary albumin-to-creatinine ratio[11]. It is noteworthy that both **PREVEND** and **Lifelines** have contributed data to CKDGen GWAS meta-analyses, either as discovery or replication cohort.

Measurement of kidney outcomes

In clinical and research practice, kidney function is assessed as glomerular filtration rate (GFR), which is the rate of pre-urine production that is obtained by filtering blood in the glomeruli. The most accurate measurements of GFR are derived from the injection of exogenous markers such as inulin, or radioisotopes such as ^{125}I-iothalamate. These markers are ideal for assessing kidney function, as their rate of excretion is dependent on their filtration through the glomerulus, and not on secretion or reabsorption in the renal tubule. However, the use of these markers for GFR measurement is costly and time-consuming, and therefore currently unavailable for large epidemiological studies. In such studies GFR is

therefore usually not measured but estimated from endogenous filtration markers that can be measured in serum, of which creatinine is the most widely used. However, given that creatinine is a product of muscle metabolism, creatinine-estimated GFR may in part reflect muscle mass rather than kidney function per se, thereby introducing bias in estimates of GFR. An alternative marker is cystatin C. This marker is not sensitive to variations in muscle mass, although other extrarenal, non-GFR factors partly explain cystatin C serum concentration. It has been shown that equations that incorporate both creatinine and cystatin C provide the most reliable estimates of GFR[29,30]. In this thesis, I therefore estimated GFR based on both creatinine and cystatin C whenever possible (**Chapter 2, 3, 5** and **7**). Furthermore, I used cystatin C estimated GFR as a positive control to creatinine-estimated GFR (**Chapter 5** and **8**). However, despite many improvements over the past decade, the accuracy of estimating equations is a debated topic[31-33]. Novel filtration markers such as beta-2-microglobulin, beta-trace-protein, and metabolite profiles, as well as the combination of these markers in novel estimating equations, may eventually result in a more accurate approximation of GFR[34,35]. This will not only lead to improved risk stratification, but also in increased power for (genetic) epidemiological studies with kidney function as trait of interest.

Albuminuria is a measure of kidney damage and a predictor of cardiovascular morbidity and mortality. Measurement of urinary albumin excretion in 24h urine collections is considered the gold standard. However, 24h collections are cumbersome, and therefore not always available in large epidemiological cohorts. A more convenient method to detect albuminuria is to measure albumin and creatinine concentrations in spot urine specimens, and then calculating the urinary albumin-to-creatinine ratio; adjusting for creatinine is a method to take into account variation in albumin concentrations due to concentration/dilution dependent on hydration status. Urinary albumin excretion and urinary albumin-to-creatinine ratio correlate well, although misclassification can occur e.g. due to differences in muscle mass[36]. Due to the poor availability of 24h urine collections for large samples, GWAS on albuminuria have thus far used outcomes based on urinary albumin-to-creatinine ratio[11,37]. A major strength of this thesis is the availability of 24h urine collections, which facilitates gold standard outcome definitions (**Chapter 2, 3, 4, 5**, and **8**), and where necessary, verification of results from urinary albumin-to-creatinine ratio based outcomes (**Chapter 4** and **8**).

Educational attainment as an indicator of socioeconomic status

Socioeconomic status is defined as the social standing or class of an individual or group. In this thesis, I used EA as the main indicator of socioeconomic status. EA is sometimes preferred because of its comparatively easy measurement and high response rate. Education is usually completed in young adulthood and predicts occupation and income, and therefore is expected to show overlap with these indicators of socioeconomic status in their association with CKD. Theoretically however, indicators of socioeconomic status are not interchangeable and have different implications. EA reflects cognitive functioning, material and intellectual resources from the family of origin, and health literacy[38-40] As EA is usually completed in young adulthood, it is unlikely that there is reverse causation by chronic diseases, such as CKD, that usually occur at later age. However some selection may be present as health at young age affects EA. A more direct measure of socioeconomic status is income. Income is a proxy for the material resources an individual can convert to health-enhancing commodities and services, and arguably the best indicator of actual, material living standards[39]. However, income may be sensitive to reporting bias, and income may not necessarily reflect disposable income, which is dependent on household composition, taxations, and hypothecated income (e.g. food stamps) that are difficult to measure. Furthermore, health may directly affect income, and therefore reverse causation may bias the results.

Recent meta-analyses that synthesize the literature on the relation between socioeconomic status and CKD[2,3] show clear associations of socioeconomic indicators, such as low EA, and low household income, with higher prevalence of CKD, lower kidney function measures, and higher levels of kidney damage markers. These meta-analyses also showed large heterogeneity between study populations, which may possibly be explained by between-country differences in lifestyle, ethnicity, educational and healthcare systems, and/or differences in risk factor prevalence. Additionally, the strength of each indicator may vary between countries. For example, it has been demonstrated that in the US, a nation with high income inequality, low income is more strongly associated with CKD than low education. In contrast to the US, health care access is less income-dependent in the Netherlands, which may explain that low income does not seem to result in excess CKD risk in the Netherlands[41]. Given that most of the data I used for this thesis were sampled from the Dutch population, I chose to use EA rather than income as the main indicator of socioeconomic status. To allow comparison with

other countries, I mapped Dutch educational levels to the International Standard Classification of EDucation (ISCED)[42].

EA was not associated with eGFR in cross-sectional analyses conditional on age (**Chapter 7**). Strong confounding with age may explain this lack of cross-sectional association: in the Netherlands, schooling until age 16 has been compulsory by law since 1969 (Dutch: "leerplichtwet"). Since 2007, due to an amendment to the 1969 law, those aged between 16-18 years can drop out only if they have a qualification (Dutch: "kwalificatieplicht") equivalent to, or higher than, secondary vocational schooling (Dutch: MBO ≥ level 2) or higher secondary schooling (Dutch: HAVO/VWO)[43] (ISCED level 3). Before 1969, schooling was only compulsory for children aged 6-14. This policy may explain that in the Netherlands, low EA (ISCED < level 3) is less prevalent among more recent cohorts (e.g. those born after the 1950s), and that those with low EA have higher age on average. In longitudinal analyses however, I observed a convincing educational gradient in eGFR decline (**Chapter 2** and **Chapter 7**) and CKD incidence (**Chapter 2**) independent of age. Future work may include a more in-depth examination of cohort effects. In particular, the cohort effects that relate to past educational policy changes may provide additional insights into the effects of EA on CKD. Potentially, if several methodological challenges can be overcome (e.g. identifying a control group, or an exogenous source of variation in exposure), these policy changes can be analyzed as natural experiments[44-46].

Heritability, GWAS, and genetic scores

Heritability is the fraction of interindividual variation of a trait that can be attributed to genetic factors, in a given population[47]. Studies that yield insights into the heritability of a disease or a trait provide clues regarding their causes, and are a first step towards disentangling genetic and environmental effects. Furthermore, heritability estimates indicate an upper bound of the proportion of phenotypic variance in traits that can be explained by genetic factors. Traditional methods for estimating heritability include the twin study, in which phenotypic similarity of identical (monozygotic) twin pairs is compared with non-identical (dizygotic) twin pairs[48]. However, twin studies potentially overestimate heritability due to unaccounted gene x gene interaction, gene x environment interaction, gene-environment correlation and violations of assumptions[49-52]. Furthermore, obtaining a representative sample of twins is difficult, and may lead to reduced

statistical power and generalizability. An alternative is to recruit families rather than twins. Through leveraging the multigenerational family design of Lifelines, I obtained heritability estimates of a number of kidney traits, including eGFRcrea, albuminuria, and serum urea (**Chapter 4**), and found that the heritability of these traits is considerable. A popular method of identifying potential genetic factors for any heritable trait is the GWAS, a data-driven, hypothesis-free method of skimming the genome for associated genetic variants. Below, I discuss some basic concepts of GWAS and discuss the methods applied in the GWAS I performed in **Chapter 6**.

The human genome consists of >6 billion nucleotide bases (guanine, cytosine, thymine, and adenine; G, C, T, A), arranged in base pairs (G-C or T-A). Most of these are fixed: any random pairing of two individuals will show >99.5% overlap in genomic sequence. The genetic factors underlying differences in traits are believed to reside within the remaining <0.5% of the genome. The most common type of variation is the single-nucleotide polymorphism (SNP), a naturally occurring variation in a single nucleotide base at specific positions in the genome, with an average frequency of ~1 in 1000 nucleotides[53]. As an example, most individuals may have a G nucleotide at a certain position (the reference allele), but in some, the position is instead occupied by an A nucleotide (the alternative allele). SNPs in coding regions potentially affect the protein product of a gene, whereas SNPs in non-coding regions may tag functional SNPs that are in linkage disequilibrium (LD, the non-random association of alleles) in coding regions, or may affect gene expression. Much of the heritability of traits may potentially be traced back to SNPs. Each individual has a paternally and a maternally inherited allele, therefore an individual can have 0 1, or 2 reference alleles of each SNP. In GWAS, these SNP alleles are tested for their association with a trait, assuming allele effects are additive. In a typical GWAS in European samples, ~10^6 independent SNP tests are performed, increasing the risk of false positive findings. To minimize this risk the consensus for genome-wide significance has been set to a strict, Bonferroni adjusted threshold of $p = 0.05/10^6 = 5 \times 10^{-8}$. A source of bias in GWAS estimates of genetic effects is population stratification: genetic drift or ancestry may lead to systematic differences in allele frequencies between subgroups in a sample. These systematic differences may lead to confounded effect estimates. Genetic principal components (PCs) may capture variation due to possible subgroup effects, and I therefore adjusted for these PCs in those studies in which I assessed SNP effects (**Chapter 5** through **8**).

In **Chapter 6**, I performed a GWAS on serum urea, a heritable indicator of kidney function. With GWAS on EA and eGFRcrea reaching sample sizes of over a million participants[10,23], the GWAS in **Chapter 6** is a relatively small study (N = 20,500). This may explain that the SNPs identified in this study only explained ~0.6% of variance in serum urea. In **Chapter 4**, I estimated the heritability of serum urea to be 30%, meaning that many of the genetic factors underlying this trait remain to be discovered. Nevertheless, the results were highly replicable and consistent with previous studies in non-European ancestry samples. The results inform studies that explore the biological functions of the identified genetic loci, and their relevance to urea metabolism and kidney function.

In addition to providing biological insights, GWAS results may be used for trait prediction. Generally, SNPs that are identified in GWAS have small effects. A genetic risk score (GRS, in this thesis also referred to as weighted genetic score, WGS) aggregates these effects, thereby greatly increasing statistical power compared to using single SNP effects. Therefore, the GRS is a practical summary score of genetic predisposition for the traits addressed in this thesis: eGFR in **Chapter 5**, **6**, **7**, and EA in **Chapter 8**. There are however limitations to the GRS. Importantly, the different genetic scores used in this thesis only explain a modest fraction of between-individual variation in traits: a 63-SNP GRS for eGFR explained only 1% in eGFR variance in PREVEND, while a 1271-SNP GRS for years of schooling explained 4% of EA in Lifelines. With ever-increasing sample sizes for GWAS, it is expected that more SNPs will eventually be detected, with effects that are estimated more precisely. Furthermore, up until now GWAS have mostly been limited to study the effects of common SNPs (i.e. SNPs with allele frequencies of ≥1%), as these could be economically genotyped with the usual GWAS arrays. As sequencing techniques become more affordable, whole genome sequencing for large samples will become feasible in the near future. With such whole genome sequence data becoming available, rarer variants (with allele frequencies well below 1%) can be detected that are predicted to have greater effects[54,55]. It is expected that these rare variants will explain a substantial part of the heritability that has thus far been hidden[52,56]. An updated GRS incorporating these rare SNPs may be a more comprehensive summary measure of genetic risk.

The GRSs used in this thesis were comprised of genome-wide significant SNPs. However, non-significant SNPs may contain additional information and thus can contribute to trait and disease prediction. Methods have been developed to

include these non-significant SNPs into genome-wide polygenic scores (PGS). For coronary artery disease, such a PGS has been reported to identify individuals with elevated risk with a predictive power comparable to that of rare monogenic mutations that typically convey a several-fold increase in disease risk[57]. Future work could include the evaluation of such PGSs based on the recent eGFR[10] and urinary albumin-to-creatinine ratio[11] GWASs for kidney outcomes.

With regards to **Chapter 7**, several specific limitations of the GRS need to be addressed. First, by using a GRS in interaction analysis, it is implicitly assumed that all genetic variants included in the GRS have directionally consistent interaction effects with EA. Another implicit assumption is that the same set of genetic variants affect eGFR in each category of EA. To check these assumptions, single SNP interaction effects need to be assessed, but this requires large sample sizes and is therefore beyond the scope of this thesis. Future research may include genome-wide interaction studies (GWIS) to identify the genetic variants whose effects are modified by EA. Similar GWIS have been performed to investigate a range of health behaviors (e.g. smoking, alcohol consumption, physical activity) and their modifying role in genetic effects on blood pressure, lipid levels, and obesity[58-61].

In each study in this thesis, I investigated European ancestry populations, and therefore I cannot generalize my findings to other ethnicities. Of note, disparities in GWAS exist, as currently, most GWAS are performed in European ancestry populations[62-64]. This is also true for the GWAS I performed in **Chapter 6**, and the GWASs that were used to create the GRSs in this thesis. Given that there may be subtle ethnic differences in the genetic architecture of disease and social traits, such as eGFR and EA, a GRS based on data from white populations may not perform similarly in populations with other ethnicities[65]. This is problematic, given that socioeconomic status is closely related to ethnicity[66-68], and that it is likely that ethnic background influences the effect of socioeconomic status on CKD[69,70]. Furthermore, if the ethnicity-gap in genetic knowledge is not bridged, this may in itself contribute to socioeconomic disparities[71]. Future work should therefore include GWAS in a multiethnic context, and expansion of GWAS into non-white populations. This allows for the creation of more inclusive and/or ethnicity-specific GRSs, and thereby allow for a more comprehensive examination of the effects of socioeconomic status on CKD.

Causal inference

In this thesis, I examined several factors for their association with kidney outcomes. Ideally, to establish causality, one would design a controlled experiment. In such a setting confounding bias would be minimized, and differences in outcomes could be attributed to intervention/exposure effects. However, experimentally establishing a causal effect of education on kidney outcomes would be unfeasible due to ethical and practical reasons: participants would have to be randomly assigned to different educational levels at a young age, and undergo follow-up for several decades until CKD occurs in mid-to-late life. Instead, the research in this thesis was based on observational data and thus, conclusions regarding causality should be interpreted with caution. Below, I describe a number of strategies employed in this thesis to strengthen causal claims in observational studies.

In **Chapters 2**, **3**, **5**, and **6**, I applied a longitudinal study design. To a certain extent, a longitudinal design helps in causal inference as it provides evidence of temporality, that is, whether the hypothesized explanatory variable precedes the outcome variable, or whether there is in fact reverse causation. In **Chapter 3**, I performed a replication study of a previous observational study to assess the consistency of the heart rate variability-CKD association, that is, whether I would reach similar conclusions regarding this association in a different, independent sample to the original discovery sample. In **Chapter 6**, replication analyses were built into the GWAS study design to assess consistency of SNP effects, thereby strengthening the conclusions in this chapter.

To minimize confounding bias, I performed multivariable analyses and adjusted my estimates for a number of known risk factors presumably influencing both exposure and outcome. However, estimates will only truly be unbiased if all confounding factors are accounted for and measured precisely, both of which are unverifiable conditions. Because of this, confounding is a threat to observational studies in general and, therefore, also a limitation of the observational studies reported in this thesis.

To examine mechanisms through which EA could affect CKD, I performed mediation analysis. Mediation refers to the mechanism in which the exposure affects the outcome (fully or partly) through a mediator variable, in which

exposure and mediator are on the same causal pathway. In **Chapter 2**, I examined several risk factors presumed to be mediators of the EA-CKD association. To estimate their mediation effects, I applied causal mediation analysis, a method within the counterfactual framework[72]. A counterfactual outcome is the potential outcome that would have occurred if the exposure were different, i.e. counter to fact; with everything else held constant, differences in the outcome can be attributed to differences in the exposure. In the mediation analysis applied in **Chapter 2**, counterfactuals of exposure and mediator variables were simulated from the original data using a bootstrap procedure. Then, from the bootstrap simulations of exposures, mediators, and outcomes, I estimated average direct effects and mediation effects. I examined mediation effects of clinical risk factors (hypertension, diabetes, high cholesterol, overweight) and health behaviors (smoking, alcohol, diet) separately, but not in conjunction with each other. This exploratory approach was chosen given that the theoretical framework regarding the interplay of these different variables is incomplete. Furthermore, time-varying effects of mediators were not considered, as the methodology to incorporate these effects has only recently been developed[73]. Future work may expand the models to include effects of multiple potential mediators, and to include time-varying effects using methods such as structural equation modelling[74].

As previously mentioned, confounding and reverse causation limit causal inference in observational studies. To strengthen causal inference in observational research, methods such as Mendelian randomization (MR) may be considered. In **Chapter 8**, I performed an MR study to assess causal effects of EA on kidney outcomes. MR is a form of instrumental variable analysis, a method applied to minimize confounding in observational studies[75]. Instrumental variables are proxies of a given exposure that must meet the exclusion restriction criterion: the instrument is related to the outcome only through the exposure. It has been proposed that individual genotype can be used as an instrumental variable[76]. Genetic variants are randomly assigned during meiosis, and therefore unrelated to any confounders. Furthermore, given that genetic variants are fixed throughout life, there cannot be reverse causation. MR studies therefore resemble an intention-to-treat analysis of a randomized controlled trial, in which participants are assigned to an intervention group based on random assignment. This randomization procedure ensures equal distribution of confounding factors in each intervention group, thus a difference in outcome between intervention groups can assumed to be due

to exposure to the intervention. A number of methods are available within the MR framework. In **Chapter 8**, I applied a two-sample MR design using summary genetic data[77], as well as a one-sample MR design using a GRS in individual-level data[78]. The potential of MR, as well as its limitations and possible threats, have been extensively described in literature[76,79-83]. Below, I discuss arguably the most important threat to MR, namely violation of the exclusion restriction criterion due to pleiotropy of genetic instruments.

MR provides unbiased causal estimates on a given exposure-outcome relation if assumptions regarding instrument validity are met. An important criterion for validity is that the genetic variant only affects the outcome through the exposure. If the biological function of a genetic variant is well-defined, this strengthens the conclusions drawn from MR. As an example, Holmes et al. used the rs1229984 variant in the alcohol dehydrogenase 1B gene (*ADH1B*) as a genetic instrument to study the effect of alcohol consumption on risk of coronary heart disease[84]. The *ADH1B* gene is known to play a specific role in alcohol metabolism, and certain variants in this gene are known to influence tolerance to alcohol and therefore consumption of alcohol. Hence, individuals are randomly assigned to alcohol consumption based on their genotype for *ADH1B*, and the effect of *ADH1B* gene variants on coronary heart disease can be attributed to alcohol exposure. For a complex trait such as EA, the biological functions of the 1271 genetic variants that have been identified in the most recent GWAS on EA[23] are poorly known. Furthermore, variants identified in GWAS may not to be causal themselves but may be linked to causal variants through linkage disequilibrium. Importantly, it is possible many of these variants have pleiotropic effects that influence risk of CKD not only through EA, but also through other pathways. Such horizontal pleiotropy may result in invalid estimates, in particular when the pleiotropy is unbalanced: unbalanced pleiotropy results in a net positive or negative bias in causal estimates. Methods have been developed that are robust to varying degrees of violations of MR assumptions due to pleiotropy, including MR Egger[85], outlier adjustment, and median- and mode based methods[86,87], and the methodology is quickly advancing. In **Chapter 8**, in case of suspected pleiotropy, I applied a range of complementary MR methods to test the robustness of my findings. In general, these complementary methods yielded results comparable to that of standard inverse variance weighted MR, thus strengthening my conclusions.

As is the case in all observational studies, including MR, selection, or conditioning on a selection variable, may lead to collider bias. A collider is a variable that is causally downstream of both exposure and outcome; conditioning on such a variable may result in biased effect estimates and spurious associations[88,89]. For the MR results in **Chapter 8**, collider bias is a possible explanation of the counterintuitive detrimental effect of EA on urinary albumin-to-creatinine ratio found in two-sample MR. However, the results of two-sample MR were corroborated in the Lifelines cohort. No selection criteria for either EA or kidney traits were applied for recruitment into Lifelines or its genotyped subset. Lifelines was assessed to be generally representative of its source population of the Northern part of the Netherlands, with only slight undersampling of those with low EA[27,90,91]. Therefore, selection bias is likely only minor and hence unlikely to have seriously affected the MR results[92]. Nevertheless, some selection is inherent given that this MR study is based on genetic data sampled from high income countries with relatively few barriers to health care. Inclusion of genetic data sampled from low to middle income countries in future studies may yield more generalizable results.

MR is a powerful method that may resolve a number of problems with causal inference from observational data, especially the problems that arise due to confounding and reverse causation. However, many potential threats (e.g. instrument pleiotropy) affect MR, and thus it is not a panacea. Some have argued that due to its many threats, null MR results are more likely to be true than non-null results[82]. Rather than above described traditional methods, MR may have a place next to these methods. Ultimately, synthesizing evidence from different designs, each with complementary sets of strengths and limitations - an approach coined 'triangulation'[93] - may be the best strategy for drawing conclusions concerning causality.

FUTURE PERSPECTIVES

Towards a better understanding of socioeconomic disparities in CKD

One of the major goals in this thesis was to elucidate the mechanisms underlying socioeconomic gradients in CKD risk, with a focus on the role of EA. Observational data have suggested that low EA is associated with CKD risk through a complex of pathways that include mediation by lifestyle factors and amplification of genetic risk. However, observational and genetic evidence from a MR study did not converge on a convincing protective effect of EA. Thus, much of the observed evidence linking EA to CKD may instead reflect an influence of other socioeconomic factors closely related to EA rather than EA per se, e.g. income, occupational factors, social deprivation, or area-level factors. Future research on socioeconomic gradients in CKD may focus on these factors rather than EA.

The data described in this thesis indicate that EA do not influence CKD risk. However, this data was sampled from high-income populations. Therefore, EA cannot be dismissed as a risk factor in lower income countries where health care access may be more dependent on EA. Inclusion of data from lower income countries, and countries where differences in EA and income are more pronounced, may yield more definitive insights into socioeconomic disparities in CKD.

Public health: opportunities for primary and secondary prevention of CKD

Given the inconsistent evidence for a protective effect of high EA against CKD in this thesis, intervention policies on EA itself, e.g. increasing school-leaving age, may not result in reduced rates of CKD. Nevertheless, low EA groups may still be a target population for preventive policies or screening. In low EA groups, the higher prevalence of modifiable renal risk factors (e.g. poor diet, smoking, high body-mass index, hypertension, and diabetes) could be a target for primary CKD prevention. Furthermore, I demonstrated that family history of CKD and a GRS based on SNPs for eGFR are associated with a higher prevalence of CKD, independent of clinical risk factors. Thus, prediction models for CKD may benefit from the inclusion of family history and/or a GRS. These models could then be used for screening purposes and early detection of CKD. Future studies may evaluate whether specifically targeting low socioeconomic status groups (defined by EA or otherwise) for primary and secondary prevention may be effective in reducing socioeconomic disparities in CKD.

Future genetic studies: bigger, more advanced, more inclusive

Contemporary genetic epidemiology is characterized by great increases in sample sizes and rapid advances in methodology, facilitated by affordable genotyping technology. These trends show no signs of slowing down, and it can be expected that due to decreasing costs, whole genome sequencing will gradually replace the usual GWAS arrays[94-96]. This means more power and more precision to identify common as well as rare genetic variants, leading to improved genetic prediction and possibly genetics-driven personalized medicine. In addition, population-based research is expected to increasingly adopt multigenerational, within-family designs, which allows for examination of, and control for, transgenerational effects[97-100]. These developments hold promise to greatly increase our understanding of the genetic underpinnings of health and behavior. Furthermore, transethnic GWAS are becoming commonplace, and an increasing number of scientists are pushing for genetic studies to be less Euro-centric and more inclusive with regards to ethnicity[62,64,71,101-103]. Thus, there is hope that the ethnicity gap in genetic knowledge will eventually be bridged, allowing a greater diversity of people to benefit from genetic data.

Collaboration

On a more general note, future studies will benefit from the continued collaboration between researchers. The disappointing results from early candidate gene studies and poor replication due to Winner's curse[104-106] has driven genetic epidemiologists to collaborate and share data on a large scale. This facilitated the inclusion of greater samples, harmonization of data, exchange of expertise, advancement of methodology, and systematic replication of results, thereby ensuring high quality, reliable science. In addition, much of the produced genetic data is made publicly available, through platforms such as the GWAS Catalog[107], LD Hub[108], and MR Base[109], which is a major stimulus for follow up study. The genetic studies performed in this thesis (**Chapter 5** through **8**) utilized data that was made possible due to such collaboration. A growing number of researchers, including those from non-genetic fields such as the social and behavioral sciences, continue to follow this example of collaboration and, by doing so, contribute to more efficient allocation of research resources and to solving the replication crisis in science[110,111].

Concluding remarks

The results in this thesis provide valuable insights into the causes of kidney disease. First, I corroborate the existence of socioeconomic disparities in kidney disease, as those with lower education tend to have higher rates of CKD and faster rates of kidney function decline. Second, those with a positive family history have a threefold higher risk of having CKD, and there is strong evidence for a genetic component to kidney traits such as eGFR, albuminuria, and serum urea. Third, genetic risk of CKD may be offset by higher socioeconomic status. Finally, educational level may not be the main driver of socioeconomic disparities in chronic kidney disease, as the genetic evidence for a causal effect of educational level is weak.

REFERENCES

1. Thio CH, Vart P, Kieneker LM, Snieder H, Gansevoort RT, Bültmann U. Educational level and risk of chronic kidney disease: Longitudinal data from the PREVEND study. *Nephrol Dial Transplant*. 2018.
2. Vart P, Gansevoort RT, Joosten MM, Bultmann U, Reijneveld SA. Socioeconomic disparities in chronic kidney disease: A systematic review and meta-analysis. *Am J Prev Med*. 2015;48(5):580-592.
3. Zeng X, Liu J, Tao S, Hong HG, Li Y, Fu P. Associations between socioeconomic status and chronic kidney disease: A meta-analysis. *J Epidemiol Community Health*. 2018;72(4):270-279.
4. Vart P, Gansevoort RT, Crews DC, Reijneveld SA, Bultmann U. Mediators of the association between low socioeconomic status and chronic kidney disease in the united states. *Am J Epidemiol*. 2015;181(6):385-396.
5. Hendriksen M, van Rossum C. Kalium inname: Risico van hyperkaliëmie?: Overzicht van beschikbare gegevens in nederland. 2015.
6. Gutierrez OM. Contextual poverty, nutrition, and chronic kidney disease. *Advances in chronic kidney disease*. 2015;22(1):31-38.
7. Thio CH, van Roon AM, Lefrandt JD, Gansevoort RT, Snieder H. Heart rate variability and its relation to chronic kidney disease: Results from the PREVEND study. *Psychosom Med*. 2018.
8. Brotman DJ, Bash LD, Qayyum R, et al. Heart rate variability predicts ESRD and CKD-related hospitalization. *J Am Soc Nephrol*. 2010;21(9):1560-1570.
9. Thio CH, van der Most, Peter J, Nolte IM, et al. Evaluation of a genetic risk score based on creatinine-estimated glomerular filtration rate and its association with kidney outcomes. *Nephrol Dial Transplant*. 2017;33(10):1757-64.
10. Wuttke M, Li Y, Li M, et al. A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat Genet*. 2019;51(6):957-972.
11. Teumer A, Li Y, Ghasemi S, et al. Genome-wide association meta-analyses and fine-mapping elucidate pathways influencing albuminuria. *Nat Commun*. 2019;10(1):1-19.
12. Pattaro C, Teumer A, Gorski M, et al. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat Commun*. 2016;7.
13. Thio CHL, Reznichenko A, van der Most PJ, et al. Genome-wide association scan of serum urea in European populations identifies two novel loci. *Am J Nephrol*. 2019;49(3):193-202.
14. Kamatani Y, Matsuda K, Okada Y, et al. Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat Genet*. 2010;42(3):210.
15. Kim YJ, Go MJ, Hu C, et al. Large-scale genome-wide association studies in east asians identify new genetic loci influencing metabolic traits. *Nat Genet*. 2011;43(10):990.
16. Okada Y, Sim X, Go MJ, et al. Meta-analysis identifies multiple loci associated with kidney function-related traits in East Asian populations. *Nat Genet*. 2012;44(8):904-909.
17. Lee J, Lee Y, Park B, Won S, Han JS, Heo NJ. Genome-wide association analysis identifies multiple loci associated with kidney disease-related traits in Korean populations. *PLOS ONE*. 2018;13(3):e0194044.
18. Prins BP, Kuchenbaecker KB, Bao Y, et al. Genome-wide analysis of health-related biomarkers in the UK household longitudinal study reveals novel associations. *Scientific reports*. 2017;7(1):11008.
19. Kirby A, Gnirke A, Jaffe DB, et al. Mutations causing medullary cystic kidney disease type 1 lie in a large VNTR in MUC1 missed by massively parallel sequencing. *Nat Genet*. 2013;45(3):299.
20. Al-Bataineh MM, Sutton TA, Hughey RP. Novel roles for mucin 1 in the kidney. *Curr Opin Nephrol Hypertens*. 2017;26(5):384-391.
21. Xu X, Eales JM, Akbarov A, et al. Molecular insights into genome-wide association studies of chronic kidney disease-defining traits. *Nat Commun*. 2018;9(1):4800.

22. van Rheenen W, Peyrot WJ, Schork AJ, Lee SH, Wray NR. Genetic correlations of polygenic disease traits: From theory to practice. *Nat Rev Genet*. 2019:1.

23. Lee JJ, Wedow R, Okbay A, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet*. 2018;50:1112-1121.

24. Tillmann T, Vaucher J, Okbay A, et al. Education and coronary heart disease: Mendelian randomisation study. *BMJ*. 2017;358:j3542.

25. Carter AR, Gill D, Davies NM, et al. Understanding the consequences of education inequality on cardiovascular disease: Mendelian randomisation study. *BMJ*. 2019;365:l1855.

26. Pinto-Sietsma SJ, Janssen WM, Hillege HL, Navis G, De Zeeuw D, De Jong PE. Urinary albumin excretion is associated with renal functional abnormalities in a nondiabetic population. *J Am Soc Nephrol*. 2000;11(10):1882-1888.

27. Scholtens S, Smidt N, Swertz MA, et al. Cohort profile: LifeLines, a three-generation cohort study and biobank. *Int J Epidemiol*. 2014.

28. Gorski M, van der Most PJ, Teumer A, et al. 1000 genomes-based meta-analysis identifies 10 novel loci for kidney function. *Scientific Reports*. 2017;7:45040.

29. Inker LA, Schmid CH, Tighiouart H, et al. Estimating glomerular filtration rate from serum creatinine and cystatin C. *N Engl J Med*. 2012;367(1):20-29.

30. Levey AS, Inker LA, Coresh J. GFR estimation: From physiology to public health. *Am J Kidney Dis*. 2014;63(5):820-834.

31. Porrini E, Ruggenenti P, Luis-Lima S, et al. Estimated GFR: Time for a critical appraisal. *Nat Rev Nephrol*. 2019;15(3):177-190.

32. Levey AS, Coresh J, Tighiouart H, Greene T, Inker LA. Strengths and limitations of estimated and measured GFR. *Nat Rev Nephrol*. 2019:1-1.

33. Porrini E, Ruggenenti P, Luis-Lima S, et al. Reply to 'Strengths and limitations of estimated and measured GFR'. *Nat Rev Nephrol*. 2019:1-2.

34. Karger AB, Inker LA, Coresh J, Levey AS, Eckfeldt JH. Novel filtration markers for GFR estimation. *EJIFCC*. 2017;28(4):277-288.

35. Steubl D, Inker LA. How best to estimate glomerular filtration rate? novel filtration markers and their application. *Curr Opin Nephrol Hypertens*. 2018;27(6):398-405.

36. Vart P, Scheven L, Heerspink HJL, et al. Urine albumin-creatinine ratio versus albumin excretion for albuminuria staging: A prospective longitudinal cohort study. *Am J Kidney Dis*. 2016;67(1):70-78.

37. Teumer A, Tin A, Sorice R, et al. Genome-wide association studies identify genetic loci associated with albuminuria in diabetes. *Diabetes*. 2016;65(3):803-817.

38. Krieger N, Williams DR, Moss NE. Measuring social class in US public health research: Concepts, methodologies, and guidelines. *Annu Rev Public Health*. 1997;18(1):341-378.

39. Galobardes B, Shaw M, Lawlor DA, Lynch JW, Davey Smith G. Indicators of socioeconomic position (part 1). *J Epidemiol Community Health*. 2006;60(1):7-12.

40. Galobardes B, Lynch J, Smith GD. Measuring socioeconomic position in health research. *Br Med Bull*. 2007;81(1):21.

41. Vart P, Gansevoort RT, Coresh J, Reijneveld SA, Bultmann U. Socioeconomic measures and CKD in the United States and the Netherlands. *Clin J Am Soc Nephrol*. 2013;8(10):1685-1693.

42. UNESCO Institute for Statistics. International standard classification of education. 2011.

43. Rijksoverheid. Leerplicht en kwalificatieplicht. www.rijksoverheid.nl/onderwerpen/leerplicht/leerplicht-en-kwalificatieplicht. Accessed Sep 7, 2019.

44. Craig P, Cooper C, Gunnell D, et al. Using natural experiments to evaluate population health interventions: New medical research council guidance. *J Epidemiol Community Health*. 2012;66(12):1182-1186.

45. Davies NM, Dickson M, Smith GD, Van Den Berg, Gerard J, Windmeijer F. The causal effects of education on health outcomes in the UK biobank. *Nat Hum Behav*. 2018;2(2):117.

46. Hamad R, Nguyen TT, Bhattacharya J, Glymour MM, Rehkopf DH. Educational attainment and cardiovascular disease in the united states: A quasi-experimental instrumental variables analysis. *PLoS Med*. 2019;16(6):e1002834.

47. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era - concepts and misconceptions. *Nat Rev Genet*. 2008;9(4):255-266.

48. Boomsma D, Busjahn A, Peltonen L. Classical twin studies and beyond. *Nature reviews genetics*. 2002;3(11):872.

49. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Nat Acad Sci*. 2012;109(4):1193-1198.

50. Purcell S. Variance components models for gene–environment interaction in twin analysis. *Twin Res Hum Genet*. 2002;5(6):554-571.

51. Felson J. What can we learn from twin studies? A comprehensive evaluation of the equal environments assumption. *Soc Sci Res*. 2014;43:184-199.

52. Nolte IM, Tropf FC, Snieder H. Missing heritability of complex traits and diseases. *eLS*. 2001:1-9.

53. International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature*. 2007;449(7164):851.

54. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461(7265):747.

55. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet*. 2010;11(6):415.

56. Wainschtein P, Jain DP, Yengo L, et al. Recovery of trait heritability from whole genome sequence data. *bioRxiv*. 2019:588020.

57. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*. 2018;50(9):1219.

58. Justice AE, Winkler TW, Feitosa MF, et al. Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits. *Nat Commun*. 2017;8:14977.

59. Sung YJ, de Las Fuentes L, Winkler TW, et al. A multi-ancestry genome-wide study incorporating gene–smoking interactions identifies multiple new loci for pulse pressure and mean arterial pressure. *Hum Mol Genet*. 2019;28(15):2615-2633.

60. de Vries PS, Brown MR, Bentley AR, et al. Multiancestry genome-wide association study of lipid levels incorporating gene-alcohol interactions. *Am J Epidemiol*. 2019;188(6):1033-1054.

61. Kilpeläinen TO, Bentley AR, Noordam R, et al. Multi-ancestry study of blood lipid levels identifies four loci interacting with physical activity. *Nat Commun*. 2019;10(1):376.

62. Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature News*. 2016;538(7624):161.

63. Liu L, Kiryluk K. Genome-wide polygenic risk predictors for kidney disease. *Nat Revi Nephrol*. 2018;14(12):723.

64. Mills MC, Rahal C. A scientometric review of genome-wide association studies. *Commun biol*. 2019;2(1):9.

65. Francisco M, Bustamante CD. Polygenic risk scores: A biased prediction? *Genome medicine*. 2018;10(1):100.

66. National Research Council, Committee on Population. *Critical perspectives on racial and ethnic differences in health in late life.* National Academies Press; 2004.

67. Braveman PA, Cubbin C, Egerter S, et al. Socioeconomic status in health research: One size does not fit all. *JAMA*. 2005;294(22):2879-2888.

68. Williams DR, Priest N, Anderson NB. Understanding associations among race, socioeconomic status, and health: Patterns and prospects. *Health Psychol*. 2016;35(4):407.

69. Patzer RE, McClellan WM. Influence of race, ethnicity and socioeconomic status on kidney disease. *Nat Rev Nephrol*. 2012;8(9):533.

70. Vart P, van Zon SKR, Gansevoort RT, Bultmann U, Reijneveld SA. SES, chronic kidney disease, and race in the U.S.: A systematic review and meta-analysis. *Am J Prev Med*. 2017;53(5):730-739.

71. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet*. 2019;51(4):584.

72. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. *Psychol Methods*. 2010;15(4):309.

73. VanderWeele TJ, Tchetgen Tchetgen EJ. Mediation analysis with time varying exposures and mediators. *J Royal Stat Soc: Series B (Statistical Methodology)*. 2017;79(3):917-938.

74. MacKinnon D. *Introduction to statistical mediation analysis*. Routledge; 2012.

75. Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol*. 2000;29(4):722-729.

76. Davey Smith G, Hemani G. Mendelian randomization: Genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet*. 2014;23(R1):R89-R98.

77. Burgess S, Scott RA, Timpson NJ, Smith GD, Thompson SG, EPIC-InterAct Consortium. Using published data in mendelian randomization: A blueprint for efficient identification of causal risk factors. *Eur J Epidemiol*. 2015;30(7):543-552.

78. Burgess S, Thompson SG. Use of allele scores as instrumental variables for mendelian randomization. *Int J Epidemiol*. 2013;42(4):1134-1144.

79. Davey Smith G, Ebrahim S. 'Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol*. 2003;32(1):1-22.

80. Smith GD, Ebrahim S. Mendelian randomization: Prospects, potentials, and limitations. *Int J Epidemiol*. 2004;33(1):30-42.

81. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stat Med*. 2008;27(8):1133-1163.

82. VanderWeele TJ, Tchetgen Tchetgen EJ, Cornelis M, Kraft P. Methodological challenges in mendelian randomization. *Epidemiology*. 2014;25(3):427-435.

83. Holmes MV, Ala-Korpela M, Smith GD. Mendelian randomization in cardiometabolic disease: Challenges in evaluating causality. *Nat Rev Cardiol*. 2017;14(10):577.

84. Holmes MV, Dale CE, Zuccolo L, et al. Association between alcohol and cardiovascular disease: Mendelian randomisation analysis based on individual participant data. *BMJ*. 2014;349:g4164.

85. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through egger regression. *Int J Epidemiol*. 2015;44(2):512-525.

86. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol*. 2016;40(4):304-314.

87. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol*. 2017;46(6):1985-1998.

88. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10:37-48.

89. Munafò MR, Tilling K, Taylor AE, Evans DM, Davey Smith G. Collider scope: When selection bias can substantially influence observed associations. *Int J Epidemiol*. 2017;47(1):226-235.

90. Klijs B, Scholtens S, Mandemakers JJ, Snieder H, Stolk RP, Smidt N. Representativeness of the LifeLines cohort study. *PloS one*. 2015;10(9):e0137203.

91. Nolte IM, van der Most, Peter J, Alizadeh BZ, et al. Missing heritability: Is the gap closing? an analysis of 32 complex traits in the lifelines cohort study. *Eur J Hum Genet*. 2017;25(7):877.

92. Gkatzionis A, Burgess S. Contextualizing selection bias in mendelian randomization: How bad is it likely to be? *Int J Epidemiol*. 2018;48(3):691-701.

93. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. *Int J Epidemiol*. 2016;45(6):1866-1886.

94. Goodwin S, McPherson JD, McCombie WR. Coming of age: Ten years of next-generation sequencing technologies. *Nat Rev Genet*. 2016;17(6):333.

95. Visscher PM, Wray NR, Zhang Q, et al. 10 years of GWAS discovery: Biology, function, and translation. *Am J Hum Genet*. 2017;101(1):5-22.

96. Wetterstrand KA. DNA sequencing costs: Data from the NHGRI genome sequencing program (GSP). www.genome.gov/sequencingcostsdata. Updated 2019. Accessed Oct 21, 2019.

97. Kong A, Thorleifsson G, Frigge ML, et al. The nature of nurture: Effects of parental genotypes. *Science*. 2018;359(6374):424-428. doi: 10.1126/science.aan6877.

98. Selzam S, Ritchie SJ, Pingault J, Reynolds CA, O'Reilly PF, Plomin R. Comparing within- and between-family polygenic score prediction. *Am J Hum Genet*. 2019;105(2):351-363.

99. Koellinger PD, de Vlaming R. Mendelian randomization: The challenge of unobserved environmental confounds. *Int J Epidemiol*. 2019;48(3):665-671.

100. Brumpton B, Sanderson E, Hartwig FP, et al. Within-family studies for mendelian randomization: Avoiding dynastic, assortative mating, and population stratification biases. bioRxiv. 2019:602516.

101. Martin AR, Teferra S, Möller M, Hoal EG, Daly MJ. The critical needs and challenges for genetic architecture studies in africa. Curr Opin Genet Dev. 2018;53:113-120.

102. Popejoy AB, Ritter DI, Crooks K, et al. The clinical imperative for inclusivity: Race, ethnicity, and ancestry (REA) in genomics. Hum Mutat. 2018;39(11):1713-1720.

103. Manolio TA. Using the data we have: Improving diversity in genomic research. Am J Hum Genet. 2019;105(2):233-236.

104. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. Nat Genet. 2001;29(3):306-309.

105. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet. 2003;33(2):177.

106. Duncan LE, Ostacher M, Ballon J. How genome-wide association studies (GWAS) made traditional candidate gene studies obsolete. Neuropsychopharmacology. 2019;44(9):1518-1523.

107. MacArthur J, Bowler E, Cerezo M, et al. The new NHGRI-EBI catalog of published genome-wide association studies (GWAS catalog). Nucleic Acids Res. 2016;45(D1):D896-D901.

108. Zheng J, Erzurumluoglu AM, Elsworth BL, et al. LD hub: A centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics. 2017;33(2):272-279.

109. Hemani G, Zheng J, Elsworth B, et al. The MR-base platform supports systematic causal inference across the human phenome. eLife. 2018;7:e34408.

110. Ioannidis JP. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124.

111. Ioannidis JP. How to make more published research true. *PLoS Med*. 2014;11(10):e1001747.

Nederlandse samenvatting
Acknowledgements
About the author
List of publications
Previous SHARE dissertations

CHAPTER 10

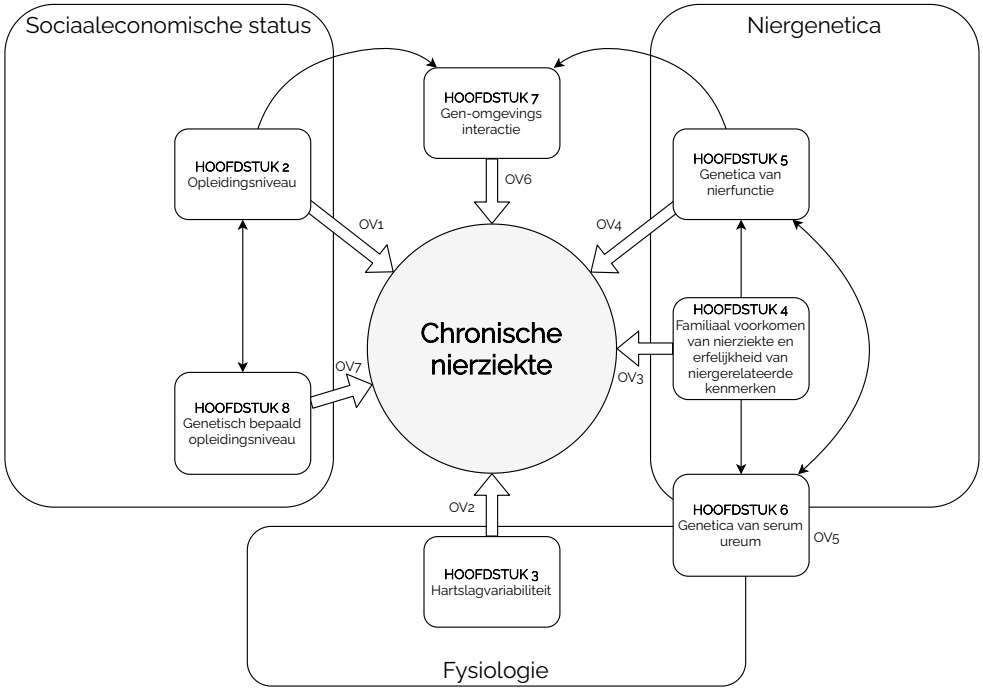
CHAPTER 10

NEDERLANDSE SAMENVATTING

INTRODUCTIE

In dit proefschrift onderzocht ik mechanismen die het risico op chronische nierziekte (CNZ) beïnvloeden. In het bijzonder was ik geïnteresseerd in sociaaleconomische ongelijkheden in CNZ, en hoe kennis van de genetische structuur van CNZ kan helpen in het begrip van deze ongelijkheden. Hiertoe verrichtte ik een aantal studies, gebruik makende van een verscheidenheid aan methodes. Naast traditionele epidemiologische methodes, zoals cohortstudies en survivalanalyse, paste ik moderne, genetisch-epidemiologische designs en concepten toe, zoals genetische risicoscores (GRS), familiestudies, de genoom-brede associatiestudie (Engels: genome-wide association study, GWAS), de gen-omgeving interactiestudie, en de Mendeliaanse randomisatie (MR) studie. In **Hoofdstuk 1** bespreek ik wat er reeds bekend is in de literatuur, en introduceer ik de onderzoeksvragen. In **Figuur 1** geef ik een grafisch overzicht van de verschillende hoofstukken, de onderzoeksvragen (OV1-7), en hun onderlinge verband in dit proefschrift.

Figuur 1. Grafische weergave van de verschillende onderzoeksvragen (OV) en hun onderlinge verband in het kader van dit proefschrift. Witte pijlen geven de hypothetische richting van het effect weer. Zwarte pijlen geven de overlap weer tussen de verschillende hoofdstukken met betrekking tot de gebruikte data, methodes, en/of resultaten.



Het onderzoek voor dit proefschrift bestaat uit een zevental studies, elk apart beschreven in **Hoofdstuk 2** t/m **8**. Het overkoepelende doel van dit onderzoek was het verkrijgen van een beter inzicht in de oorzaken van CNZ, en in het bijzonder in de sociaaleconomische verschillen in het voorkomen van CNZ. Thematisch kan onderscheid gemaakt worden tussen traditioneel (sociaal-)epidemiologisch onderzoek (**DEEL I: Hoofdstuk 2** en **3**), genetisch-epidemiologisch onderzoek (**DEEL II: Hoofdstuk 4**, **5** en **6**), en een combinatie van beide (**DEEL III: Hoofdstuk 7** en **8**). Hieronder volgt een samenvatting van mijn bevindingen.

DEEL I: EVALUATIE VAN HET EFFECT VAN SOCIAALECONOMISCHE STATUS EN AUTONOME DYSFUNCTIE OP HET RISICO OP CHRONISCHE NIERZIEKTE

Hoofdstuk 2: Opleidingsniveau is geassocieerd met risico op chronische nierziekte in de algemene bevolking

Dat wat bekend is over de relatie tussen sociaaleconomische status en CNZ is grotendeels gebaseerd op cross-sectioneel onderzoek (d.w.z. onderzoek gebaseerd op een dwarsdoorsnede van de bevolking) uit de Verenigde Staten. In **Hoofdstuk 2** beschrijf ik een longitudinaal onderzoek (d.w.z. onderzoek waarbij mensen gedurende een bepaalde tijd worden gevolgd) dat ik heb uitgevoerd met data van de PREVEND studie, een Nederlandse observationele studie met deelnemers die zijn gerekruteerd uit de algemene bevolking. In dit onderzoek gebruikte ik het hoogst behaalde opleidingsniveau als maat voor sociaaleconomische status. Ik laat zien dat, vergeleken met deelnemers met een hoog opleidingsniveau (HBO of hoger), diegenen met een laag opleidingsniveau (MBO niveau 2 of lager) een 25% hoger risico hebben op het ontwikkelen van CNZ. Daarnaast hebben diegenen met een laag opleidingsniveau een grofweg 15% per jaar snellere nierfunctieachteruitgang hebben. Verkennende analyses suggereren dat dit wordt veroorzaakt door het vaker voorkomen van hoge bloeddruk, suikerziekte, en overgewicht bij diegenen met een laag opleidingsniveau.

De resultaten verkregen uit deze studie dragen in een belangrijke mate bij aan de bestaande literatuur vanwege de verkregen inzichten in de onderliggende mechanismen van de relatie tussen opleidingsniveau en CKD.

Een belangrijke bevinding van deze studie is dat dieet mogelijk een mediërende rol speelt: er is suggestief bewijs dat laag opleidingsniveau tot CNZ leidt via een

ongezond dieet, en in het bijzonder lage kaliuminname (d.w.z. weinig groente en fruit). Ondanks dat een dergelijke mediërende rol reeds lang werd verondersteld, was er tot nu toe weinig bewijs hiervoor.

Een andere belangrijke bevinding is dat hoge bloeddruk, suikerziekte, overgewicht, en een ongezond dieet niet volledig verklaren waarom lager opgeleiden een hoger risico hebben op nierfunctieachteruitgang. Dit suggereert dat er andere, niet-traditionele risicofactoren zijn die de relatie tussen opleidingsniveau en CNZ mediëren. Toekomstig onderzoek zou zich kunnen richten op het identificeren van deze, nog onbekende, factoren. Verder zou kunnen worden onderzocht hoe de verschillende mediërende factoren zich tot elkaar verhouden. Met een completer beeld van de mechanismen die bij laagopgeleiden leiden tot CNZ, zouden we een rangschikking kunnen maken van factoren waarop kan worden ingegrepen teneinde sociaaleconomische ongelijkheden in CNZ te verkleinen.

Hoofdstuk 3: Een lage hartslagvariabiliteit gaat niet vooraf aan chronische nierziekte in de algemene bevolking

Niet-traditionele risicofactoren kunnen een rol spelen in de ontwikkeling van CNZ. Een voorbeeld van een dergelijke factor is een lage hartslagvariabiliteit (Engels: heart rate variability, HRV) als maat voor een dysfunctioneel autonoom zenuwstelsel. In **Hoofdstuk 3** beschrijf ik een onderzoek waarin ik onderzocht of een lage HRV invloed heeft op het ontwikkelen van CNZ.

Een lage HRV komt voor als er een dysfunctie is van het autonome zenuwstelsel, waarbij de functie van de parasympathicus gereduceerd is ten opzichte van die van de sympathicus. Mogelijk leidt een autonome dysfunctie tot nierschade via veranderingen in de renale bloedsomloop; bewijs voor een dergelijk effect komt voort uit dierexperimenteel onderzoek. Mogelijk geldt dit ook voor mensen: een eerder onderzoek in de VS leverde observationeel bewijs voor een associatie van een lage HRV met CNZ-gerelateerde ziekenhuisopname en nierfalen. De verwachting was daarom dat een lage HRV ook voorafgaat aan nieuw ontstane CNZ.

Ik verrichtte een longitudinale analyse in data van 4605 deelnemers aan de PREVEND studie. Hier observeerde ik een 50-100% hoger risico op CNZ in diegenen in het laagste kwartiel (d.w.z. de laagste 25%) van HRV ten opzichte

van die van de bovenste drie kwartielen. Echter, dit sterk verhoogde risico leek volledig te worden verklaard door een hogere leeftijd van diegenen met een lage HRV. Om deze reden kon ik een verband tussen een lage HRV en CNZ niet bevestigen. In verkennende analyses vond ik bewijs voor een omgekeerd effect: omdat een lage HRV alleen geassocieerd was met verminderde nierfunctie in diegenen die reeds CNZ hadden, is het waarschijnlijker dat CNZ leidt tot een lage HRV in plaats van andersom.

DEEL II: DE GENETICA VAN NIERFUNCTIE EN DE VERTALING NAAR DE KLINIEK EN NAAR ONDERZOEK

Hoofdstuk 4: De invloed van erfelijkheid op nier-gerelateerde kenmerken is aanzienlijk, en een positieve familieanamnese is een belangrijke determinant van chronische nierziekte in de algemene bevolking

Lifelines is een unieke studie vanwege zijn familiedesign: meerdere generaties van meerdere families zijn gerekruteerd uit de algemene bevolking. In deze data maakte ik een schatting van de erfelijkheid, d.w.z. de bijdrage van genetische factoren aan interindividuele verschillen, van een aantal nier-gerelateerde kenmerken. Ik observeerde aanzienlijke erfelijkheidsschattingen van onder meer geschatte glomerulaire filtratiesnelheid (Engels: estimated glomerular filtration rate, eGFR; 44%), urine albumine (20%) en serum ureum (31%). Het risico op het hebben van CNZ, in het geval van een ziek eerstegraadsfamilielid, is drie keer hoger dan het risico in de algemene bevolking. Deze studie, beschreven in **Hoofdstuk 4**, is de grootste familiestudie naar nier-gerelateerde kenmerken en CNZ tot nu toe. Een belangrijke bevinding is dat een positieve familieanamnese het risico op CNZ sterk verhoogt. Verder markeren de erfelijkheidsschattingen een bovengrens voor de hoeveelheid variatie in nier-gerelateerde kenmerken die kan worden verklaard door genetische factoren. Toekomstige studies zouden zich kunnen richten op het identificeren van deze factoren.

Hoofdstuk 5: Een genetische risicoscore gebaseerd op 53 eGFRcrea SNPs is een bruikbare genetische proxy van nierfunctie, maar mogelijk niet van gevoeligheid voor chronische nierziekte

Naast traditionele risicofactoren en levensstijl spelen genetische factoren een rol in CNZ. Genoom-brede associatiestudies (Engels: genome-wide association studies, GWAS) hebben genetische varianten ontdekt die geassocieerd zijn met CNZ definiërende maten, namelijk eGFR en urine albumine. Tot nu toe zijn de

GWAS naar eGFR het meest succesvol gebleken, gelet op het aantal ontdekte varianten. Ten tijde van het schrijven van **Hoofdstuk 5** waren er 53 enkel-nucleotide polymorfismen (Engels: single nucleotide polymorphisms, SNPs) bekend die associaties vertoonden met eGFR geschat op basis van serum creatinine (eGFRcrea). Per SNP zijn de effecten zeer klein en daarom niet betekenisvol in de kliniek. Het is echter mogelijk om alle SNP effecten samen te voegen tot één verzamelscore, de zogenaamde genetische risicoscore (GRS). Een dergelijke GRS zou in de kliniek kunnen worden gebruikt om onderscheid te maken tussen degenen met hoger en lager risico voor CNZ. Ook zou deze GRS kunnen worden gebruikt voor onderzoeksdoeleinden, bijvoorbeeld als maat voor genetische aanleg voor CNZ. In dit hoofdstuk evalueerde ik de bruikbaarheid van een GRS gebaseerd op eerdergenoemde 53 SNPs geassocieerd met eGFRcrea. In 3649 deelnemers aan de PREVEND studie vond ik bescheiden maar robuuste associaties met eGFRcrea. Deze resultaten valideerde ik vervolgens met een op cystatine C gebaseerde schatting van eGFR (eGFRcysc) als uitkomst: ik vond vergelijkbare associaties van de GRS met eGFRcysc. Dit is van belang aangezien eGFRcrea een onnauwkeurige maat is voor nierfunctie vanwege de relatie tussen creatinine en spiermassa, een relatie die voor cystatine C niet geldt. Een andere belangrijke bevinding was dat de GRS niet geassocieerd was met urine albumine, en een associatie met eGFR vertoonde ongeacht de traditionele risicofactoren voor CNZ, namelijk body-mass index, roken, hypertensie, diabetes, hoog cholesterol, en een voorgeschiedenis van cardiovasculaire ziekte. Dit geeft aan dat de GRS daadwerkelijk een genetische proxy is van nierfunctie, en niet van nierschade of risicofactoren voor CNZ. Echter werd slechts 1% van de variatie in eGFR verklaard door de GRS. Daarbovenop waren de resultaten van longitudinale analyse niet eenduidig: de GRS was niet geassocieerd met nierfunctieachteruitgang, en vertoonde los van baseline eGFR geen associaties met incidente CNZ. Mijn conclusie is daarom dat een GRS, gebaseerd op onze huidige kennis van de genetica van nierfunctie, geen rol van belang heeft in de kliniek in bijvoorbeeld individuele ziektevoorspellingen. Echter, omdat de GRS een ware genetische proxy is van nierfunctie, zou de GRS kunnen worden gebruikt in onderzoek op populatieniveau, en in studiedesigns zoals Mendeliaanse randomisatie.

Hoofdstuk 6: Een genoom-brede associatiestudie van serum ureum in Europeanen identificeert nieuwe associaties met twee genetische loci

In **Hoofdstuk 5** gebruikte ik twee verschillende diagnostische markers voor eGFR,

namelijk serum creatinine en serum cystatine C. In **Hoofdstuk 6** onderzocht ik de genetica van serum ureum, een andere diagnostische marker van nierfunctie waarvan ik de erfelijkheid aantoonde in **Hoofdstuk 4**. Genetisch onderzoek naar deze marker is voornamelijk gedaan in Oost-Aziaten, waarbij SNPs in of nabij 13 genetische loci bekend waren ten tijde van het schrijven van dit hoofdstuk. Slechts enkele genetische studies hebben deze marker onderzocht in Europeanen, waarbij vijf associaties werden gevonden. Echter, deze vijf associaties waren tot dan toe nog niet gerepliceerd. Daarom voerde ik de eerste meta-analyse van GWAS uit op serum ureum in populaties van Europese komaf, met een ontdekkingsfase in 13.312 deelnemers aan de Lifelines studie, en een replicatiefase in 7379 deelnemers aan drie cohorten uit de algemene bevolking (PREVEND, NESDA, en EGCUT). Ik identificeerde repliceerbare associaties van zes SNPs die genoom-breed significant waren ($p < 5x10^{-8}$). Van deze zes SNPs waren er twee in of nabij genetische loci die niet eerder in verband waren gebracht met serum ureum, namelijk rs2003313 op chromosoom 11 nabij *POU2AF1*, en rs998394 op chromosoom 3 in *ADAMTS9-AS2*. Verder kon ik alle SNPs repliceren die eerder in Oost-Aziaten en Europeanen waren gevonden, op SNPs in of nabij drie loci na die mogelijk specifiek gelden voor Oost-Aziaten.

Ik onderzocht verder de overlap van mijn bevindingen met genetische data over eGFRcrea. Enige overlap tussen serum ureum en eGFRcrea kan worden verwacht, aangezien de serumwaardes van beide worden beïnvloed door nierfunctie. Eerder werd een hoge genetische correlatie van r_g=0.56 gevonden tussen ureum en creatinine. De positieve richting van deze genetische correlatie geeft aan dat ureum en creatinine genetische factoren delen, en dat deze genetische factoren de serumwaarden van beide in de zelfde richting beïnvloeden, oftewel hogere serum ureum is genetisch gecorreleerd met hogere serum creatinine (en dus lagere eGFRcrea). Verder bewijs voor genetische overlap tussen deze twee markers lever ik door aan te tonen dat de 53 eGFRcrea SNPs (dezelfde 53 SNPs als in **Hoofdstuk 5**) verrijkt waren voor serum ureum: 14 van de 53 eGFRcrea SNPs waren ook geassocieerd met serum ureum, veel meer dan kan worden verwachten op basis van kans. Een GRS op basis van deze 53 SNPs vertoonde bescheiden maar significante associaties met serum ureum. Na correctie voor eGFRcrea verzwakte de associatie van de GRS met serum ureum, wat suggereert dat de GRS een effect heeft op serum ureum via nierfunctie.

Ondanks de statistische significantie is de klinische waarde van mijn bevindingen beperkt: de geïdentificeerde SNPs verklaarden samen niet meer dan 0.56% van de variatie in serum ureum. Mijn bevindingen genereren echter hypotheses voor twee nieuwe genetische loci (*POU2AF1* en *ADAMTS9-AS2*) en hun relatie tot serum ureum en nierfunctie, die kunnen worden getest in functioneel onderzoek. Verder kunnen de GWAS resultaten op serum ureum worden gebruikt voor het valideren van mogelijke nierfunctie loci: als een bepaalde genetische variant inderdaad een marker is voor nierfunctie, kan worden verwacht dat deze variant zowel gerelateerd is met hoger eGFRcrea als met lager serum ureum.

DEEL III: HET BENUTTEN VAN GENETICA BIJ HET VERKLAREN VAN SOCIAALECONOMISCHE ONGELIJKHEDEN IN CHRONISCHE NIERZIEKTE

Hoofdstuk 7: Laag opleidingsniveau versterkt het genetische risico op chronische nierziekte in de algemene bevolking

In **Hoofdstukken 7** en **8** gebruikte ik de kennis die ik opdeed in de voorgaande hoofdstukken om genetische methodes te integreren met traditionele sociaal-epidemiologische methodes. In **Hoofdstuk 7** presenteer ik bewijs, verkregen uit de PREVEND studie, voor een versterkend effect van laag opleidingsniveau op het genetische risico op verminderde nierfunctie. Deze bevinding was het meest uitgesproken in longitudinale analyse, waar ik een interactie vond tussen opleidingsniveau en GRS: een snellere nierfunctieachteruitgang werd gezien in diegenen met een hogere GRS en een lager opleidingsniveau. Dit effect was groter dan de som van de effecten van de GRS en opleidingsniveau. Mijn bevindingen suggereren daarnaast dat een hoog opleidingsniveau het genetische risico op verminderde nierfunctieachteruitgang teniet doet, aangezien in deze groep geen effect van de GRS kon worden aangetoond. De interactie kon slechts voor een beperkt deel worden verklaard door traditionele risicofactoren voor nierziekte, wat suggereert dat er niet-gemeten factoren bestaan waarvan de invloed niet wordt gevangen door traditionele factoren.

Deze resultaten zijn een belangrijke bijdrage aan de literatuur: het zijn de eerste die bewijs leveren voor een gen-omgevingsinteractie in nierfunctie met een modificerend effect van opleidingsniveau. Een belangrijke nevenbevinding is dat de genetische aanleg voor nierfunctie gelijk verdeeld is over verschillende opleidingsniveaus; er is geen bewijs voor selectie voor risicovarianten in

diegenen met een laag opleidingsniveau. Daarom kan het hogere risico op nierfunctieachteruitgang in diegenen met een laag opleidingsniveau worden toegeschreven aan een versterking van genetische aanleg, door een omgeving die samenhangt met een lage opleiding. Deze resultaten pleiten tegen genetisch determinisme, d.w.z. het ontwikkelen van verminderde nierfunctie is niet vooraf gebaseerd op basis van je genen. Gezien het zeer bescheiden interactie-effect (verklaarde variantie 0.1% in nierfunctieachteruitgang) is de klinische toepasbaarheid met betrekking tot individuele risicovoorspelling beperkt. Verder zal uit toekomstig onderzoek moeten blijken of het effect reproduceerbaar is in andere populaties. In dat geval zou er op populatieniveau voordeel kunnen worden verwacht van interventies die ingrijpen op laag opleidingsniveau, en de modificeerbare gevolgen hiervan, in het verkleinen van sociaaleconomische ongelijkheden in nierfunctieachteruitgang. Hoewel de hier beschreven studie genoeg statistische power had om een interactie-effect te vinden met continue uitkomsten, zijn grotere aantallen deelnemers nodig om te bepalen of dit interactie-effect te vinden is met CNZ als uitkomst. Tot slot rechtvaardigen de resultaten verder onderzoek naar de mediërende paden tussen laag opleidingsniveau en CNZ, alsook naar de specifieke genen die betrokken zijn in deze paden.

Hoofdstuk 8: De associatie tussen opleidingsniveau en chronische nierziekte is mogelijk niet oorzakelijk

De resultaten uit **Hoofdstuk 2** suggereren dat een hoger opleidingsniveau een beschermend effect heeft op de nieren, gegeven dat in die studie een hoger opleidingsniveau was geassocieerd met langzamere nierfunctieachteruitgang en lagere incidentie van CNZ. Het is echter onzeker of dit een ware oorzakelijke relatie betreft gezien de observationele aard van het bewijs. In **Hoofdstuk 8** pas ik een Mendeliaanse randomisatie (MR) methode toe, waarin ik gebruik maak van genetische instrumenten voor opleidingsniveau. Door het gebruik van genetische instrumenten wordt het risico op bias beperkt. Hierdoor kunnen, met meer zekerheid dan met traditionele observationele methodes, conclusies worden getrokken met betrekking tot causaliteit. Voor een twee-sample MR analyse verwierf ik gegevens over 1271 SNPs waarvan een bekende associatie is met het aantal voltooide schooljaren. Vervolgens extraheerde ik de associaties van deze SNPs uit genetische data van het CKDGen Consortium aangaande eGFRcysc, eGFRcrea, en albuminurie (urine albumine-creatinine ratio). Ik vond dat elke standaard deviatie (4.2 jaar) hoger opleidingsniveau was geassocieerd

met een 3.2% hogere eGFRcysc, conform een beschermende effect van opleidingsniveau op nierfunctie. Voor eGFRcrea vond ik echter geen effect van opleidingsniveau. Hoger opleidingsniveau was geassocieerd met een hogere urine albumine-creatinine ratio, wat suggereert dat hoger opleidingsniveau resulteert in nierschade. Deze contra-intuïtieve bevinding onderzocht ik verder in individuele data van deelnemers aan de Lifelines studie, een studie die beschikt over gedetailleerde urine albumine data. Ik construeerde een genetische score die bestond uit eerdergenoemde 1271 SNPs, en gebruikte deze score als genetisch instrument voor aantal voltooide schooljaren in een één-sample MR analyse. Het in twee-sample MR gevonden contra-intuïtieve nadelige effect van hoger opleidingsniveau op urine albumine-creatinine ratio werd ook gevonden in Lifelines. Door de beschikbaarheid van 24-uurs urine in Lifelines kon ik bevestigen dat er inderdaad sprake was van verhoogde urine albumine uitscheiding en niet verminderde urine creatinine uitscheiding (d.w.z. niet een gevolg van vertroebeling door spiermassa). De resultaten suggereren dat hoger opleidingsniveau leidt tot hogere albuminurie.

Er is een grote verzameling aan bewijs voor een beschermend effect van opleidingsniveau op cardiovasculaire gezondheid; ik verwachtte daarom ook een beschermend effect op renale gezondheid. De effecten van opleidingsniveau op eGFRcrea en eGFRcysc waren echter niet eenduidig, en ik vond een onverwacht nadelig effect op albuminurie. Toekomstig werk zou zich kunnen richten op het verklaren van deze onverwachte bevinding. Op basis van deze resultaten concludeer ik dat er onvoldoende genetisch bewijs is voor een beschermend causaal effect van opleidingsniveau op renale gezondheid. Toekomstige studies naar ongelijkheden in CNZ zouden zich kunnen richten op andere mogelijk oorzakelijke sociaaleconomische factoren zoals inkomen, beroep, beroepsmatige blootstellingen, sociale achterstand, of factoren op buurtniveau.

ALGEMENE DISCUSSIE

In **Hoofdstuk 9** geef ik een samenvatting van de voornaamste bevindingen en bespreek ik de context waarin deze kunnen worden geplaatst. Daarnaast komen de gebruikte methodes uitgebreid aan bod: ik bespreek de gebruikte data, de manier waarop nierfunctie en sociaaleconomische status is gemeten, de genetische methodes, en zet ik uiteen welke strategieën ik heb gebruikt om tot causale gevolgtrekkingen te komen. Op basis van de bevindingen en gebruikte methodes

geef ik aanbevelingen voor vervolgonderzoek naar de sociaaleconomische en genetische determinanten van CNZ. Tot slot deel ik mijn visie over hoe het beter kan, en waar de sociale en genetische epidemiologie naar toe gaat.

CONCLUSIES

Het onderzoek in dit proefschrift is een combinatie van sociaal- en genetisch epidemiologisch onderzoek in de nefrologie. De resultaten leveren belangrijke inzichten in de oorzaken van chronische nierziekte. Ten eerste bevestig ik dat er sociaaleconomische verschillen zijn in het ontwikkelen van chronische nierziekte, gezien de bevinding dat mensen met een lager opleidingsniveau vatbaarder zijn voor het ontwikkelen van chronische nierziekte en een sterkere nierfunctieachteruitgang vertonen. Ten tweede hebben mensen met een positieve familieanamnese een drie keer verhoogd risico op chronisch nierziekte, en is er sterk bewijs voor een genetische component van nier-gerelateerde kenmerken zoals eGFR, albuminurie, en serum ureum. Ten derde zou een hogere sociaaleconomische status het genetische risico op chronische nierziekte teniet kunnen doen. Tenslotte is het mogelijk dat opleidingsniveau niet de belangrijkste aandrijver is van sociaaleconomische verschillen in chronisch nierziekte, gezien het ontbreken van eenduidig genetisch bewijs hiervoor.

Acknowledgements

CHAPTER 10

ACKNOWLEDGEMENTS/DANKWOORD

Dat was het dan! We zijn aangekomen bij het einde van mijn PhD traject. Echter, niemand komt ver zonder een beetje geluk en de hulp van begeleiders, collega's, vrienden, en familie.

Allereerst wil ik mijn promotieteam bedanken: Ron Gansevoort, Ute Bültmann, en Harold Snieder. Dit proefschrift is het resultaat van een verregaande samenwerking tussen verschillende disciplines, en jullie hebben dat makkelijk doen lijken. Ik hoop nog veel met jullie te werken. Ron, bedankt voor je enthousiasme, toewijding, humor, en kritische blik. Het is absurd hoe snel (en op welke tijdstippen!) je stukken van terugkoppeling voorziet, ik weet niet waar je de tijd vandaan haalt. In het begin botsten we wat (niet in de minste plaats door mijn eigenwijsheid), maar onze discussies tilden het werk altijd naar een hoger niveau. Ute, ik denk met plezier terug aan onze meetings; zeker de meetings die we één-op-één hadden brachten me inzichten over het werk, niet alleen inhoudelijk maar ook eromheen. Deze inzichten zijn onmisbaar geweest voor dit proefschrift en voor mijn persoonlijke groei. Mocht Sigur Rós weer in het land zijn pikken we het concert mee! Beste Harold, soms zit het mee en tref je de juiste persoon, op de juiste plaats, en op het juiste moment. Jij bent die persoon; je zag de wetenschapper in mij en bood jezelf aan als mentor. Het heeft mij de vereiste focus gebracht en mij definitief op het onderzoekspad gezet. Bedankt voor al je vertrouwen in mij!

I would very much like to thank the reading committee, Philipp Koellinger, Nine Knoers, and Anna Köttgen, for assessing this dissertation and approving it for a defense.

I am grateful for having been a part of the Unit of Genetic Epidemiology and Bioinformatics for the past few years. It has been a warm and safe place ever since I joined it, due to the laid back and collaborative athmosphere, and the many brilliant (ex-)members that helped me in one way or the other: Adolfo, Anna L, Ali Reza, Azmeraw, Anna L, Bale, Behrooz, Bram, Fahimeh, Jana, Judith, Katri, Leanne, Loretto, Marzyeh, Nigus, Peter T, Renske, Rujia, Shifteh, Siqi, Sophie, Tian, Valeria, Xueling, Yldau, Zhen, thank you! Ilja, heel erg bedankt voor alle hulp met betrekking tot statistiek en programmeren die ik van je heb gekregen de afgelopen jaren. Anna R, thank for you laying the groundwork for the serum urea GWAS and for the

informal chit chat during the few conferences we both attended (I fondly remember Regensburg). Jia, in the short time you were in Groningen you managed to leave a great impression. Thank you for the great collaboration on the heritability paper and I hope more joint papers will follow! Ahmad, we were office buddies during the last leg of your PhD, and I had the honor of being your paranymph. Thank you for the wise lessons (don't work too hard!) and your valuable bioinformatics expertise. Zoha, thank you for running the bioinformatics pipeline for the serum urea GWAS. I'm quite impressed by your work and look forward to all you will achieve!

Special thanks go out to my paranymphs, fellow Unit members and also former office buddies during the critical years of my PhD research: Peter van der Most, Eliza Fraszczyk, and Anna Neustaeter. Thank you for putting up with my loud sighs, grunts and curses; for helping me out with bugged R scripts; for the history lectures; for all the drinks, decorated desks and birthday gifts; for sharing stories, frustrations, and happiness. I feel nostalgic for the time we spent together in our little office on the 2nd floor of the Triadegebouw. Thanks for being my friends and paranymphs. You are awesome!

Veel hulp heb ik gekregen vanuit de afdeling Epidemiologie. Hans, ondanks alle drukte was je nooit te beroerd om me te helpen als ik bij je aanklopte met mijn statistische vragen, dank daarvoor! Noha, technically I only started working with you after I finished my PhD. Nevertheless, I would like to thank you for your enthusiasm, experience, and moral support, which are quickly proving invaluable for my development as a teacher. Aukje, Anita, Roelian, en Lisette: dank voor alle logistieke hulp! Eva en Marloes: bedankt voor de fijne samenwerking op alle GECKO, PACE, en LifeCycle projecten. To all PIs, supporting staff, postdocs, and PhD students in the department: thank you all for contributing to a stimulating environment for researchers to thrive in!

Ik wil Jan Aalberts, Maarten van den Berg, en Jan Peter van Tintelen bedanken voor de wetenschapsstage die ik jaren geleden liep in de Marfan-kliniek van het UMCG. Ik heb veel bij jullie geleerd, en de vonk van het onderzoek doen is toen overgeslagen.

Mijn dank gaat uit naar de vele Nefronerds, de (ex-)nierpromovendi van de afdeling Nefrologie, Pathologie, en Farmacologie. Akin, Amber, Antonio, Bart, Claire, Elise, Irina, Isidor, Judith, Laura, Lyanne, Lianne, Maarten, Marco, Marieke, Marjolein, Marloes, Maryse, Michèle, Michelle, Skander, Suzanne, Yusof, en alle anderen: Ik ben dankbaar voor jullie gezelschap tijdens en rondom alle congressen, en de altijd levendige discussies bij de Kidney Center Meetings.

Van de afdeling Health Sciences van het UMCG bedank ik Sander en Priya. Sander, dank voor je bijdrage op het interactiepaper, je bent een fijn iemand om mee samen te werken, en ik hoop dat nog vaker te doen! Priya, thank you for all your input. Your thesis on socioeconomic disparities in CKD was a great springboard for my own work. Adelita, ontzettend bedankt dat je me bij de oren door de cursus Health Psychology heen hebt gesleurd. Ik raak echter nog steeds van streek als ik een gedragsmodel zie.

Arie en Joop, bedankt voor de hulp bij het HRV stuk! Ik ben uiteindelijk een andere richting ingegaan maar ik hoop in de toekomst weer eens met jullie te werken.

A number of genetics consortia contributed data to this thesis. I would like to thank the SSGAC for kindly providing the educational GWAS data, the ICBP for providing GWAS data on blood pressure traits, and the CKDGen Consortium for the kidney traits GWAS data and for having me as an analyst for current projects. Furthermore, I would like to thank the investigators, analysts, and participants of all study cohorts that contributed data to this thesis; these include the PREVEND study, the Lifelines Cohort and Biobank, NESDA, and EGCUT. Pim and Niek, thank you for your help with the PREVEND GWAS data.

Igor, dankjewel voor de layout van dit boekje (sorry, BOEK!) en het spijt me voor al die vreselijke tabellen die je erin hebt moeten proppen. Ik kijk met plezier terug op onze layoutsessies. Senne, dank voor je schitterende illustraties!

Om als PhD student mentaal gezond te blijven zijn vrienden en vaste ontspanningsplekken onmisbaar. Remy, Lisette, Eltjo, Kas, Bart en Susanne, Marie en Tjerk: dank voor alle pubquizzen, escape rooms, filmavondjes, festivals, concertjes, etentjes, terrasjes, potjes klaverjas, of welke afleiding dan ook. Patrick, Arjen, fijn dat we, ondanks de afstand, nog steeds met elkaar afspreken. Alle staf, vrijwilligers, en de bandjes-scene in VERA, in het bijzonder die bij Downstage, Zienema, en Beheer, bedankt! Het waren

13 mooie jaren als vrijwilliger. Vanwege een druk onderzoeksleven kom ik er niet meer regelmatig, maar het voelt nog altijd als thuis. Mark, Neil, Jan Anco en Saskia, Edwin, Douwe, Robert, Laurah, Marinke, Thomas, Marc en Simone, dank voor de toffe tijd in de kelder en daarbuiten. Dank aan iedereen bij De Minnaar, De Koffer, en De Gans: jullie weten wie jullie zijn. Adriaan, dank voor de nerd-discussies tijdens onze lunches en borrels, we gaan de crap uit die MR cursus lesgeven. Ymkje Anna, wat ben je briljant en toch ook benaderbaar. Dank dat ik altijd bij je terecht kan voor een informeel babbeltje, morele steun, en een goede lunch. Annelene en Margriet, dank dat ik bij jullie kon aansluiten tijdens de studie en tijdens het begin van mijn PhD project. Lieve zeergeleerde opponent dr. Fleur, het klikte met jou meteen en van dat enkele jaartje als CPE studie- en stadsgenoten hebben we het meeste gemaakt. Elke fatsoenlijke kans om je weer te zien grijp ik aan (zelfs in mijn oppositie!).

Bedankt Bert, Gabry, Tara, Koen, Henny en Pieter: jullie zorgen ervoor dat ik me in Scheemda helemaal thuisvoel.

Ik ben ontzettend dankbaar naar mijn lieve familie. Mijn broers Hans, Paul en Willy (en natuurlijk ook Marieke, Leoniek, Sarah, en de kinderen): dank voor jullie onvoorwaardelijke steun, zonder jullie was dit nooit gelukt. Jullie verdienen ook een verontschuldiging: af en toe leek ik van de planeet te zijn verdwenen vanwege het drukke werk. Mam, Pap, ik ben eindelijk afgestudeerd. Bedankt voor jullie liefde, toewijding, harde werk, en jullie geduld, zeker toen ik richting in mijn leven miste. Ik hou van jullie.

Lieve Esther: de laatste paragraaf is voor jou. Je hebt het hele traject van dichtbij meegemaakt, en daarmee ook mijn beste en slechtste karaktereigenschappen ervaren. Dankjewel voor je geduld, je steun, je opoffering, en je begrip. Bedankt dat je me tegenhoudt teveel te werken. Bedankt voor je bereidheid te lachen om mijn flauwe grappen. Sorry dat ik geen ochtendmens ben. Overal is het leuker met jou, en samen kunnen we alles aan. Deze mijlpaal is nu behaald, en nu zijn de avonden en weekenden weer voor ons. Ik hol van die!

Tot slot, een laatste gedachte:

Nunc est bibendum!

CHAPTER 10

ABOUT THE AUTHOR

Christian Han Liong (Chris) Thio was born on November 28th, 1983, in Hoogeveen, the Netherlands. During his Bachelor's studies in Medicine at the University of Groningen, he joined Prof Harold Snieder's Unit of Genetic Epidemiology and Bioinformatics as an intern. After obtaining his Bachelor's degree in 2013, he enrolled in the Research Master programme Clinical and Psychosocial Epidemiology. In 2015, he obtained his Research Master's degree. He then succesfully applied for funding for a PhD project of his own design, which he started that same year. The results of this project are now summarized in this PhD dissertation. For this project, he became a full-time member of the Unit of Genetic Epidemiology. Whilst pursuing his Master's and PhD degrees, he became interested in socioeconomic disparities in health and disease, in particular with regards to kidney disease and cardiometabolic traits. Furthermore, he became interested in biostatistics, (epi-)genetic epidemiology, and causal inference. After finishing his dissertation in 2019, he accepted a position of junior lecturer for the Department of Epidemiology at the Universitiy Medical Center Groningen. He is involved in teaching several courses on medical statistics and epidemiology at the undergraduate and graduate levels. Furthermore, he contributes to several international consortia-based collaborative efforts as a researcher/analyst.

List of publications

CHAPTER 10

LIST OF PUBLICATIONS

Thio CHL, van Zon SKR, van der Most PJ, Snieder H, Bültmann U, Gansevoort RT. The association of genetic factors, educational attainment, and their interaction, with kidney function outcomes. *Am J Epidemiol. In press*

Thio CHL, Reznichenko AR, van der Most PJ, Kamali Z, Vaez A, Smit JH, Penninx BWJH, Haller T, Mihailov E, Metspalu A, Damman J, de Borst MH, van der Harst P, Verweij N, Navis GJ, Gansevoort RT, Nolte IM, Snieder H; Lifelines Cohort Study group. Genome-wide association scan of serum urea in European populations identifies two novel loci. *Am J Nephrol*. 2019;49(3):193-202.

Thio CHL, van Roon AM, Lefrandt JD, Gansevoort RT, Snieder H. Heart rate variability and its relation to chronic kidney disease: Results from the PREVEND study. *Psychosom Med*. 2018;80(3):307–16.

Thio CHL, Vart P, Kieneker LM, Snieder H, Gansevoort RT, Bültmann U. Educational level and risk of chronic kidney disease: Longitudinal data from the PREVEND study. *Nephrol Dial Transplant*. 2018. *In press*

Thio CHL, van der Most PJ, Nolte IM, van der Harst P, Bültmann U, Gansevoort RT, Snieder H. Evaluation of a genetic risk score based on creatinine-estimated glomerular filtration rate and its association with kidney outcomes. *Nephrol Dial Transplant.* 2017;33(10):1757-64.

Özyilmaz A, de Jong PE, Bakker SJL, Visser ST, **Thio CHL**, Gansevoort RT; PREVEND Study Group. Screening for elevated albuminuria and subsequently hypertension identifies subjects in which treatment may be warranted to prevent renal function decline. *Nephrol Dial Transplant*. 2016;32(suppl_2):ii200-ii208.

Munoz ML, van Roon AM, Riese H, **Thio CHL**, Oostenbroek E, Westrik I, de Geus EJ, Gansevoort RT, Lefrandt JD, Nolte IM, Snieder H. Validity of (ultra-)short recordings for heart rate variability measurements. *PLoS ONE*. 2015;10(9):e0138921.

Aalberts JJ, **Thio CHL**, Schuurman AG, van Langen IM, van der Pol BA, van Tintelen JP, van den Berg MP. Diagnostic yield in adults screened at the Marfan outpatient clinic using the 1996 and 2010 Ghent nosologies. *Am J Med Genet Part A*. 2012;158(5):982-88.

Consortium contributions

Wuttke M, Li Y, Li M, et al. A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat Genet*. 2019;51(6):957-72.

Tin A, Marten J, Kuhns VLH, et al. Target genes, variants, tissues and transcriptional pathways influencing human serum urate levels. Nat Genet. 2019;51(10):1459-74.
Teumer A, Li Y, Ghasemi S, et al. Genome-wide association meta-analyses and fine-mapping elucidate pathways influencing albuminuria. *Nat Commun*. 2019;10(1):4130.

Noordam R, Young WJ, Salman R, et al. Effects of calcium, magnesium, and potassium concentrations on ventricular repolarization in unselected individuals. *J Am Coll Cardiol*. 2019;73(24):3118-31.

CHAPTER 10

Research Institute SHARE

This thesis is published within the **Research Institute SHARE** (Science in Healthy Ageing and healthcaRE) of the University Medical Center Groningen / University of Groningen. Further information regarding the institute and its research can be obtained from our internet site: www.share.umcg.nl

More recent theses can be found in the list below. (supervisors are between brackets)

2020

Raven D

Where's the need? The use of specialist mental health services in adolescence and young adulthood
(prof AJ Oldehinkel, prof RA Schoevers, dr F Jörg)

Stoter IK

Staying on track; the road to elite performance in 1500 m speed skating
(dr MT Elferink-Gemser, prof C Visscher, prof FJ Hettinga)

Silva Lagos LA

Gestational diabetes mellitus and fetoplacental vasculature alterations; exploring the role of adenosine kinase in endothelial (dys)function
(dr MM Faas, prof T Plösch, prof P de Vos, prof L Sobrevia)

Reints R

On the design and evaluation of adjustable footwear for the prevention of diabetic foot ulcers
(prof ir GJ Verkerke, porf K Postema, dr JM Hijmans)

Bekhuis E

A body-mind map; epidemiological and clinical aspects of the relation between somatic, depressive and anxiety symptomatology
(prof JGM Rosmalen, prof RA Schroevers, dr L Boschloo)

Havinga PJ

Breaking the cycle? Intergenerational transmission of depression/anxiety and opportunaties for intervention
(prof RA Schoevers, prof CA Hartman, dr L Boschloo)

Geer SJ van der

Trismus in head and neck cancer patients
(prof PU Dijkstra, prof JLN Rodenburg, dr H Reintsema)

Salavati N

Preconception environmental factors and placental morphometry in relation to pregnancy outcome
(prof JJHM Erwich, prof RM van der Beek, dr MK Bakker, dr SJ Gordijn)

Fels IMJ van der

Movement, cognition and underlying brain functioning in children
(dr E Hartman, prof C Visscher, prof RJ Bosker, dr J Smith)

Braaksma P
Moving matters for children with developmental coordination
(prof R Dekker, prof CK van der Sluis, dr MM Schoemaker, dr I Stuive)

Akkerman M
Functioning beyong pediatric burns
(prof LHV van der Woude, dr LJ Mouton, dr MK Nieuwenhuis)

Buurke TJW
Adaptive control of dynamic balance in human walking
(dr CJC Lamoth, prof LHV van der Woude, dr AR den Otter)

Bos GJFJ
Physical activity and physical fitness in children with chronic conditions
(prof PU Dijkstra, prof JHB Geertzen)

Heugten P van
Talent in international business defined: implications and applications for honours education
(prof ADC Jaarsma, dr MVC Wolfensberger, dr M Heijne-Pennninga)

Kranenburg HA
Adverse effects following cervical manual physical therapy techniques
(prof CP van der Schans, dr MA Schmidt, dr GJR Luijckx)

Lameijer CM
Perspectives on outcome following hand and wrist injury in non-osteoporotic patients
(prof CK van der Sluis, prof HJ ten Duis, dr M El Moumni)

2019
Nijholt W
Gaining insight in factors associated with successful ageing: body composition, nutrition, and cognition
(prof CP van der Schans, dr JSM Hobbelen, dr H Jager-Wittenaar)

Schaap FD
Dementia Care Mapping in the care for elderly people with an intellectual disability and dementia
(prof SA Reijneveld, dr EJ Finnema, dr GJ Dijkstra)

Willems-Jongsma M
Lifestyle change in adults with intellectual disabilities; use and effectiveness of behaviour change techniques
(prof CP van der Schans, dr A Waninge, dr J de Jong, dr TIM Hilgenkamp)

For earlier theses visit our website.

NOTES