Chronic kidney disease

Thio, C. H. L.

DOI:
10.33612/diss.133648108

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

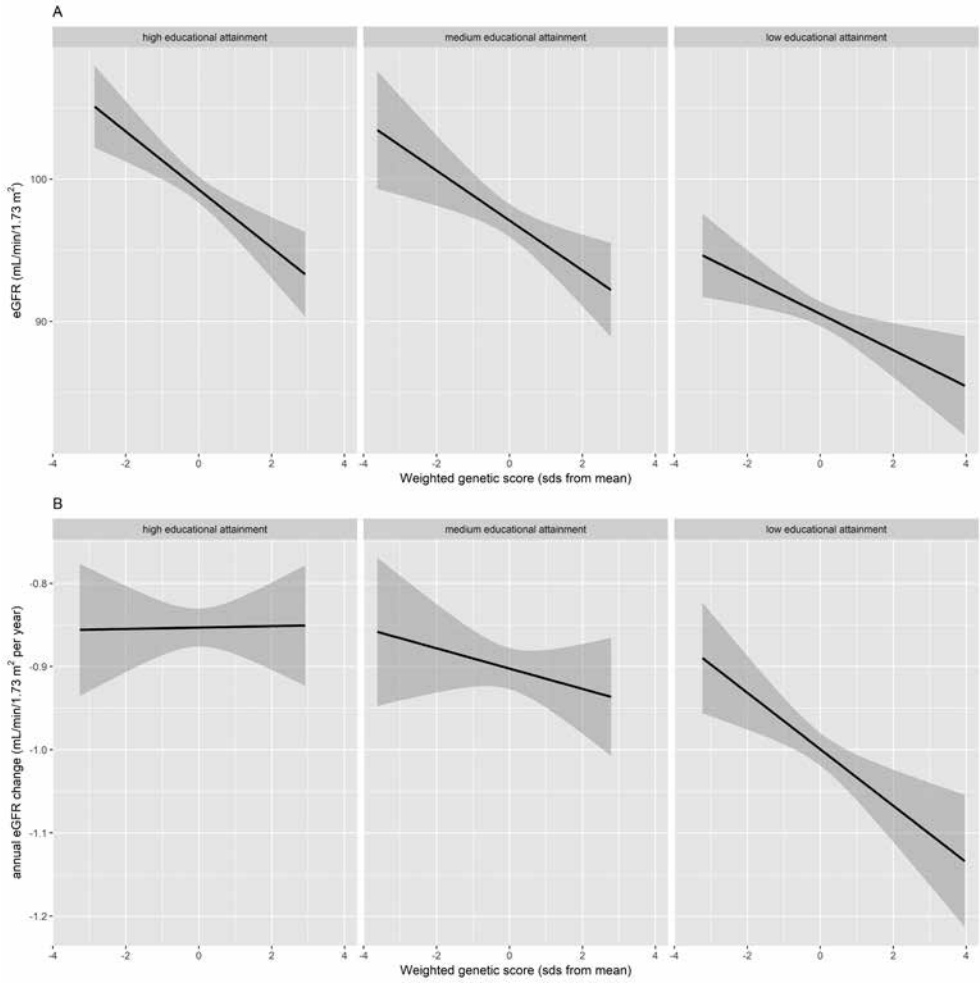
A plot of baseline eGFR by the WGS and strata of EA is presented in **Figure 1A**. On visual inspection of this data, the effect of the WGS on eGFR appeared to be consistent across strata of EA, hence, we anticipated that the interaction term between the WGS and EA in our models would not be significant. 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

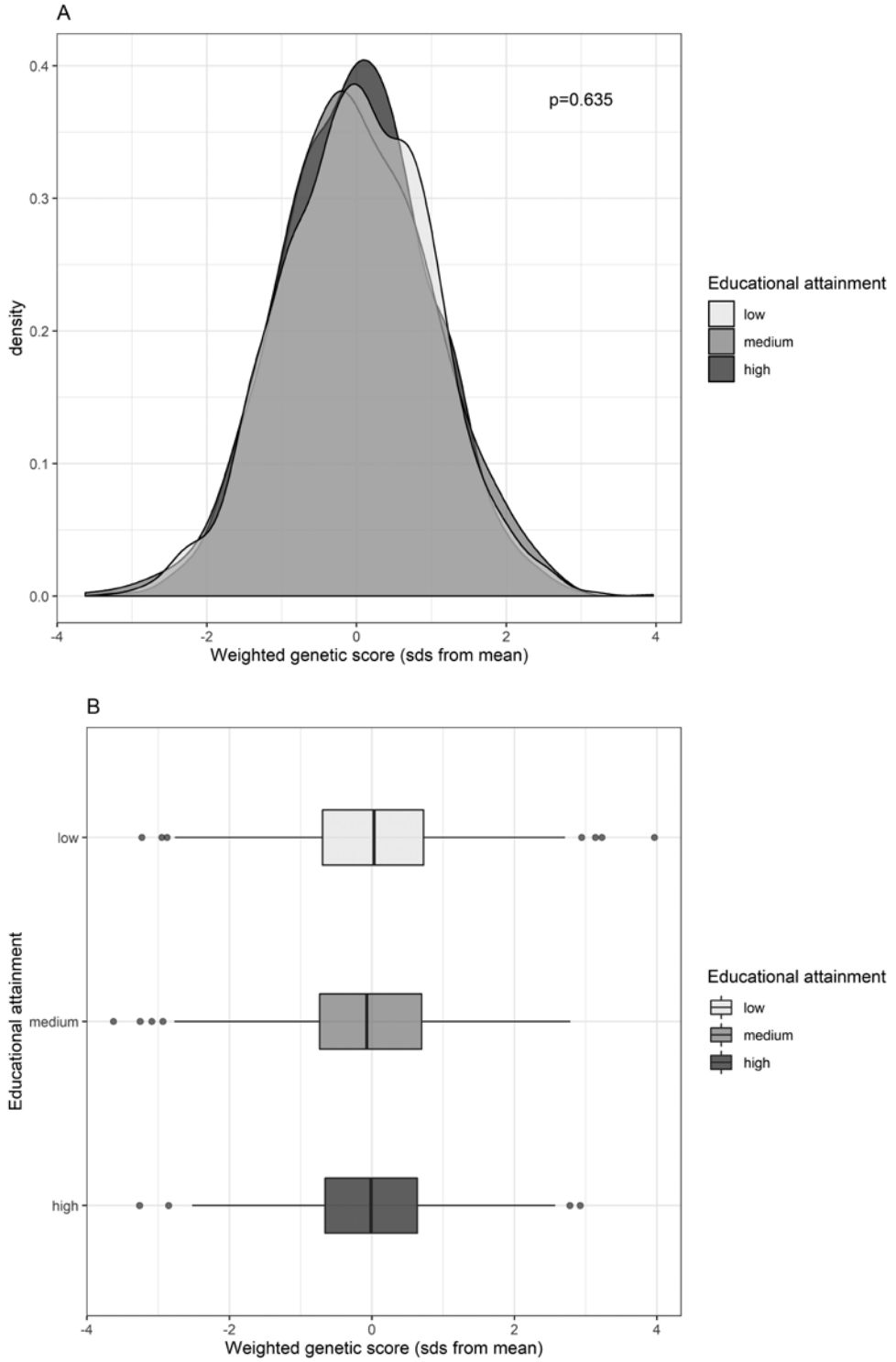
Model	df	AIC	BIC	logLik	deviance	χ^2	Δdf	p
Model 1	12	87311	87400	-43644	87287			
Model 2	16	87311	87430	-43640	87279	8.1169	4	0.087
Model 3	25	85221	85406	-42585	85171			
Model 4	29	85218	85433	-42580	85160	10.306	4	0.036
Model 5	31	82072	82301	-41005	82010			
Model 6	35	82070	82328	-41000	82000	10.698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