

University of Groningen

The association of genetic factors, educational attainment, and their interaction, with kidney function outcomes

Thio, Chris H L; van Zon, Sander K R; van der Most, Peter J; Snieder, Harold; Bültmann, Ute; Gansevoort, Ron T

Published in:
 American Journal of Epidemiology

DOI:
[10.1093/aje/kwaa237](https://doi.org/10.1093/aje/kwaa237)

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

Document Version
 Final author's version (accepted by publisher, after peer review)

Publication date:
 2020

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

Citation for published version (APA):

Thio, C. H. L., van Zon, S. K. R., van der Most, P. J., Snieder, H., Bültmann, U., & Gansevoort, R. T. (2020). The association of genetic factors, educational attainment, and their interaction, with kidney function outcomes. *American Journal of Epidemiology*, [kwaa237]. <https://doi.org/10.1093/aje/kwaa237>

Copyright

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

Take-down policy

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

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

**The association of genetic factors, educational attainment, and their interaction, with
kidney function outcomes**

Chris H.L. Thio, Sander K.R. van Zon, Peter J. van der Most, Harold Snieder, Ute Bültmann,
and Ron T. Gansevoort

Correspondence to: Chris H.L. Thio Unit of Genetic Epidemiology & Bioinformatics Dept.
of Epidemiology (HPC FA40) University Medical Center Groningen, University of
Groningen Hanzeplein 1, PO Box 30.001, 9700RB Groningen, the Netherlands (email:
c.h.l.thio@umcg.nl)

Author affiliations: University Medical Center Groningen (UMCG), University of Groningen
(RuG), Dept of Epidemiology, Unit of Genetic Epidemiology & Bioinformatics, Groningen,
the Netherlands, (Chris H.L. Thio, Peter J. van der Most, Harold Snieder); UMCG, RuG, Dept
of Health Sciences, Community and Occupational Medicine, Groningen, the Netherlands
(Sander K.R. van Zon, Ute Bültmann); UMCG, RuG, Dept of Internal Medicine, Division of
Nephrology, Groningen, the Netherlands (Ron T. Gansevoort)

© The Author(s) 2020. Published by Oxford University Press on behalf of the Johns Hopkins
Bloomberg School of Public Health. This is an Open Access article distributed under the terms
of the Creative Commons Attribution Non-Commercial License
(<http://creativecommons.org/licenses/by-nc/4.0>), which permits non-commercial re-use,
distribution, and reproduction in any medium, provided the original work is properly cited. For
commercial re-use, please contact journalpermissions@oup.com.

Funding: The PREVEND Study in general was funded by the Dutch Kidney Foundation (grant E.033).

Running head: Gene x education interaction in kidney function

ABSTRACT

Both genetic predisposition and low educational attainment (EA) are associated with higher risk of chronic kidney disease. We examined the interaction of EA and genetic risk in kidney function outcomes. We included 3,597 participants from the Prevention of Renal and Vascular End stage Disease Cohort Study, a longitudinal study in a community-based sample from Groningen, the Netherlands (median follow-up 11 years, 1997-2012). Kidney function was approximated by estimating glomerular filtration rate (eGFR) from serum creatinine and cystatin C. Individual longitudinal linear eGFR trajectories were derived from linear mixed models. Genotype data on 63 single nucleotide polymorphisms, with known associations to eGFR, were used to calculate an allele-weighted genetic score (WGS). EA was categorized into high, medium, and low. In ordinary least squares analysis, higher WGS and lower EA showed additive effects on reduced baseline eGFR; the interaction term was non-significant. In analysis of eGFR decline, the significant interaction term suggested amplification of genetic risk by low EA. Adjustment for known renal risk factors did not affect our results. This study presents the first evidence of gene-environment interaction between EA and a WGS on eGFR decline, and provides population-level insights into the mechanisms underlying socioeconomic disparities in chronic kidney disease.

Keywords: Educational attainment, genetic risk, interaction, kidney function, chronic kidney disease

Abbreviations: **BMI:** body mass index, **CKD:** chronic kidney disease, **EA:** educational attainment, **eGFR:** estimated glomerular filtration rate, **GWAS:** genome-wide association study, **PREVEND:** Prevention of RENal and Vascular ENd stage Disease **WGS:** weighted genetic score, **SBP:** systolic blood pressure, **SNP:** single nucleotide polymorphism, **UAE:** urinary albumin excretion

Chronic kidney disease (CKD) is a heterogeneous group of disorders characterized by sustained kidney dysfunction and/or signs of kidney damage¹. CKD is associated with cardiovascular morbidity and all-cause mortality². It may eventually progress to end-stage kidney disease, necessitating the start of renal replacement therapy. The incidence of CKD is increasing, posing a major global health challenge³⁻⁵.

Over the past two decades, evidence has accumulated for a socioeconomic gradient in CKD: low educational attainment (EA), as an indicator of low socioeconomic status, 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 socioeconomic status 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 CKD¹⁰⁻¹³, 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 (eGFR_{crea})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 CKD¹⁵⁻¹⁷.

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 purported downstream effects of low education to improve kidney outcomes. The joint associations 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 have been described elsewhere²². 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 used the subset of participants that was

genotyped (n=3649). Given that participants may receive education into their 20s, we excluded those aged <30 years (N=52) from the analyses, resulting in N=3,597.

Measurements

Kidney function

Kidney function was approximated by estimated glomerular filtration rate from creatinine and cystatin 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), 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) equation²⁴. Outliers exceeding four standard deviations (sds) from the mean were excluded.

Genotyping and genetic risk score calculation

Genotyping details for PREVEND were described previously¹⁷. Briefly, genotyping was performed on the Illumina CytoSNP12 v2 chip. Samples with call rate <95%, duplicates, and sex discrepancies were excluded. Markers with call rate >95%, Hardy-Weinberg equilibrium $p \geq 10^{-5}$ and minor allele frequency $\geq 1\%$ were included. Variants were imputed to 1000G Phase 1 version 3, using Minimac software. To account for population stratification, principal component analysis was performed²⁵; the resulting principal components represent possible

population substructures in PREVEND. In order to remove ethnic outliers, samples with $z\text{-score} > 3$ for any of the first five principal components with the highest eigen values were excluded. From the resulting GWAS data, we extracted genotypes of 63 known eGFR SNPs identified in a meta-analysis of GWAS on eGFR_{crea} in European populations¹⁴. We constructed a weighted genetic score (WGS) comprising these SNP. 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 by the population standard deviation. Effect alleles were those reported to associate with lower eGFR, thus a higher WGS reflects genetic predisposition towards lower kidney function.

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 Education²⁶. We then categorized EA into low (no, primary, basic vocational, and secondary education, corresponding to International Standard Classification of Education levels 0-2), medium (senior secondary vocational and general senior secondary education, International Standard Classification of Education levels 3-4), and high (higher professional and higher academic education, International Standard Classification of Education levels 5-6). International Standard Classification of Education levels were imputed to US years of schooling. High EA was the reference category in all analyses.

Covariates

We adjusted for age, age², and sex. To minimize potential confounding by population stratification, we additionally adjusted for the first ten genetic principal components. In longitudinal analyses, we additionally adjusted for baseline eGFR. Furthermore, we explored models that include the renal risk factors, body-mass-index (BMI, weight/height²), systolic blood pressure (SBP), glucose, total cholesterol, and smoking status (never smoker, former

smoker, current smoker), each measured at baseline. Furthermore, we adjusted for natural log-transformed urinary albumin excretion (*ln*UAE), an indicator of kidney damage measured in two 24h urine collections at baseline. In sensitivity analyses, we adjusted for hypertension and diabetes rather than SBP and glucose. For continuous variables, outliers exceeding four SDs from the mean were excluded.

Statistical analyses

All analyses were performed using R software (R Foundation for Statistical Computing, Vienna, Austria) version 3.5.1²⁷.

To assess the explained variance of eGFR by the WGS, conditional on age, age², sex, and the first ten principal components, $\Delta R^2_{\text{adjusted}}$ 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 regression analysis using the *lm()* function 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 implemented in the *lme4* R package²⁸, 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 regression analysis. For both cross-sectional analyses and longitudinal analyses, ten models were constructed with the main effects of the WGS and EA (models 1-10), in addition their interaction term, and varying degrees of covariate adjustment (see **WEB TABLE 1** 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 ($\Delta R^2_{\text{adjusted}}$) 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. A significant P-value for the interaction term indicates departure from additivity. Finally, EA-stratified models, with varying degrees of covariate adjustment, were constructed (models 11-15). For all models, we performed complete-case analysis. We applied a two-sided significance threshold of $\alpha=0.05$ 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).

We regressed baseline eGFR on the WGS to obtain a crude association. The association of the WGS with baseline eGFR was modest but highly significant (B (se) = -1.68 (0.29), $R^2_{\text{adjusted}}=0.010$, $P=8.6 \times 10^{-9}$).

In **WEB FIGURE 1**, we plot WGS distribution by categories of EA. The WGS was normally and equally distributed in each EA category. 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 1**. On visual inspection of this data, the association of the WGS with 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**). A one-SD increase in the WGS was associated with 1.61 mL/min/1.73m² 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*, P=0.512, **TABLE 3**). Adjustment for covariates (*models 3-4*; age, age², sex, and the first 10 principal components) 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. Inclusion of additional covariates (*models 5-8*) did not change our conclusions, although counterintuitively, low EA was significantly associated with higher eGFR in these models. The association of the WGS with baseline eGFR appeared smaller in the low EA stratum (**FIGURE 2**), but the interaction was non-significant for all models.

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/1.73m² per year (SD=0.385). A plot of eGFR change by the WGS and strata of EA is presented in **FIGURE 1**. In this figure, the WGS is shown to have its strongest association with eGFR change in those with low EA (**FIGURE 1C**). In those with medium or high EA (**FIGURE 1A-B**), the WGS had no

apparent association with eGFR change. A trend in mean eGFR change was observed across EA levels, with those with lower EA having 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/m² 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, age², sex, and the first 10 genetic principal components) increased the association of the WGS with eGFR change (*model 3*, B (se) = -0.027 (0.006), P=2.3 x10⁻⁵), while attenuating the association of EA with 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 the expected direction (*model 4*, low vs high EA, B (se) = -0.036 (0.015), P=0.017), suggesting that the joint association of the WGS and EA is greater than the sum of their main associations. The contribution of the overall interaction term between the WGS and EA was modest but significant (*model 4 vs model 3*, P=0.036, **TABLE 3**).

The influence of potential mediators (i.e. BMI, SBP, glucose, total cholesterol, and smoking status) on 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 association 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*, P=0.062, **TABLE 3**), 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). Adjustment for lnUAE did not affect our results (*model 7-8*). The WGS most strongly associated with annual eGFR change in the low EA stratum (**FIGURE 2**).

Sensitivity analysis

The WGS did not show significantly different distributions between categories of EA. However, **FIGURE 1** and **WEB FIGURE 1** 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 (eGFR_{crea}), and from serum cystatin C only (eGFR_{cysc}). Results were generally consistent with our main analysis, with EA being more strongly associated with eGFR_{cysc} than with eGFR_{crea}. Similarly, interaction effects between the WGS and EA were more pronounced for eGFR_{cysc} than for eGFR_{crea} (data not shown).

We repeated the interaction analyses using a linear mixed model only. Here, despite some minor discrepancy with longitudinal estimates from ordinary least squares regression analysis, effect estimates were generally and directionally consistent with the ordinary least squares analysis (**WEB TABLE 2**), and a three-way interaction term to assess the modifying effect of EA on WGS in eGFR change (WGS x EA x time) was again significant (**WEB TABLE 3**).

Adjustment for hypertension and diabetes, rather than SBP and glucose, did not affect our results (*model 9-10*, **TABLE 3**, **WEB TABLE 1-3**).

DISCUSSION

In the present study, we investigated the associations of genetic factors (summarized by a weighted genetic score, WGS) and educational attainment (EA), as well as the interaction between the WGS and EA, with 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 interaction suggests an amplifying effect of low EA on genetic risk, and could not 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 lower eGFR 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 disproportionately fast eGFR decline in the most vulnerable in terms of EA and genetic predisposition. Low EA is unlikely to directly amplify genetic risk of reduced eGFR. Rather, it may act through a range of interrelated purported 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 not explained by traditional renal risk factors. Therefore, other factors likely exist that explain the interaction between EA and a WGS. These may include factors with socioeconomic gradients such as health literacy²⁹, occupational exposures and infections³⁰, whose influence may not be captured by traditional risk factors.

In cross-sectional analyses, low EA was significantly associated with higher eGFR (*models 5-10*), suggesting a paradoxical protective effect of low EA on kidney function. However, this is likely the result of overadjustment bias³¹ in these models, given that many of the covariates (e.g. hypertension and diabetes) are purported mediators in the relation with EA and eGFR^{8,9}.

Each of the 63 SNPs that were identified in previous GWAS on eGFR_{crea}¹⁴ have small effect-sizes. The WGS aggregates these SNPs, thereby greatly increasing statistical power

compared to using single SNP associations. 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. Sample sizes and thus power for GWAS on eGFR have recently greatly increased, facilitating the detection of over 200 additional genetic variants³². Using a more comprehensive WGS that includes these variants likely increases power to detect interactions. In addition, participants with an equal WGS may have different underlying risk variants. 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 associations with kidney function are modified by EA. Similar studies were performed for blood pressure, BMI and lipids, for modification by smoking, alcohol use and physical activity³³⁻³⁶.

For the longitudinal analyses, we reported results from a two-step method in which we used individual eGFR trajectories, extracted from a linear mixed model, as outcome variable in ordinary least squares regression analysis. 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 associations with eGFR change compared to the present study, but did not reach statistical significance in linear mixed model analysis¹⁷. Alternatively, the associations of the

WGS, EA, and the WGS x EA interaction term on eGFR change can also be modelled in a single linear mixed model, taking into account the random variation and correlation between eGFR measurements. However, R^2 estimation is not straightforward in linear mixed models, and estimation of the interaction effect 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 a linear mixed model only. Notwithstanding some discrepancies with the ordinary least squares analysis regarding effect size and statistical significance, the results from linear mixed model were directionally consistent with ordinary least squares analysis and therefore our conclusions remain unchanged.

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 exclusively of participants of European ancestry, 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; replicability and generalizability of these results to other populations is uncertain and therefore require validation in independent samples. Under similar parameters, the interaction effect could be replicated with a sample size of ~5000 (with 80% power at $\alpha=0.05$) (**WEB FIGURE 2**). Third, the observational nature of this study precludes causal conclusions. Fourth, larger

samples are needed to examine whether the interaction between a WGS and low EA results in increased rates of CKD. Finally, a higher attrition rate was observed in those with low education. This may have resulted in bias towards the null, or underestimation of effect-sizes, 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 effect-sizes. Furthermore, given the population-based sample, our findings may not translate into the clinic, i.e. in predicting disease progression in CKD patients. 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 of genetic risk, and EA, have independent associations with the rate of kidney function decline. Furthermore, our results suggest a subtle amplifying effect of low EA on genetic risk of reduced eGFR. Traditional kidney risk factors that are purported 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, warranting further investigation.

ACKNOWLEDGEMENTS

Author affiliations: Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands (*Chris H.L. Thio, Peter J. van der Most, Harold Snieder*) ; Department of Health Sciences, Community and Occupational Medicine, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands (*Sander K.R. van Zon, Ute Bültmann*) ; Department of Nephrology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands (*Ron T. Gansevoort*)

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.

Author contributions: *Chris H.L. Thio, Sander K.R. van Zon, and Harold Snieder* designed the study. *Chris H.L. Thio* analyzed the data. *Peter J. van der Most* contributed to the data analysis. *Chris H.L. Thio* drafted the work. *Chris H.L. Thio, Sander K.R. van Zon, Peter J. van der Most, Ute Bültmann, Harold Snieder* and *Ron T. Gansevoort* contributed to the interpretation of data, as well as critical revision and final approval of the final manuscript.

Conflict of interest: none declared

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. *American Journal of Kidney Diseases.* 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.* 2020;35(7):1211-1218.
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'Seaghda 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.

Nephrology Dialysis Transplantation. 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 ISCED 2011*. United Nations Educational, Scientific and Cultural Organization, Institute for Statistics, Montreal, Canada. 2012 (UIS/2012/INS/10/REV)
27. R Core Team. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. URL <http://www.R-project.org/>. 2014.
28. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*. 2014. Accessed 22 Jan 2020.
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 American 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 International Supplements*. 2017;7(2):88-106.
31. Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology*. 2009;20(4):488-495.
32. 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.
33. 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. *Nature communications*. 2017;8:14977.

34. 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.

35. 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.

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

ORIGINAL UNEDITED MANUSCRIPT

TABLE 1. Baseline characteristics overall and by educational attainment in PREVEND (1997-2012)

Characteristic	Educational attainment							
	Total (3597)		Low (1673)		Medium (889)		High (1035)	
	Mean (SD)	%	Mean (SD)	%	Mean (SD)	%	Mean (SD)	%
Age (years)	50 [40-60] ^a		55 [46-65] ^a		46 [37-56] ^a		44 [37-51] ^a	
Males		52		49		56		53
eGFR (mL/min/1.73m ²)	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)		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)	
Hypertension		35		46		31		21
Glucose (mmol/L)	4.8 (0.8)		5.0 (0.8)		4.7 (0.7)		4.6 (0.6)	
Type 2 diabetes		4.0		5.7		3.6		1.7
BMI ^b	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] ^a		9.9 [4.2-11.6] ^a		11.1 [4.8-12.2] ^a		11.2 [6.2-12.4] ^a	

Abbreviations are: eGFR, estimated glomerular filtration rate; WGS, weighted genetic risk score; SBP, systolic blood pressure; BMI, body-mass-index.

^a Values are expressed as median [interquartile range]

^b Weight (kg)/height (m)²

TABLE 2. Results of interaction analysis from ordinary least squares regression analysis in PREVEND (1997-2012)^a

Parameter	Model 1			Model 2			Model 3			Model 4		
	B	se	p	B	se	p	B	se	P	B	se	P
eGFR, mL/min/1.73m ²												
Intercept	99.27	0.52	0	99.27	0.52	0	91.39	0.56	0	91.39	0.56	0
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
Educational attainment												
low	-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
medium	-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	0.91
high	0			0			0			0		
WGS * Educational attainment												
WGS * low				0.77	0.69	0.265				0.60	0.53	0.256
WGS * medium				0.29	0.77	0.711				0.29	0.60	0.628
WGS * high				0						0		
annual eGFR change, mL/min/1.73m ² per year ^b												
Intercept	-1.089	0.041	0	-1.090	0.041	0	-0.695	0.048	0	-0.697	0.048	0
WGS (per sd)	-0.016	0.007	0.014	0.004	0.013	0.746	-0.027	0.006	2.3 x10 ⁻⁵	-0.008	0.012	0.52
Educational attainment												
low	-0.125	0.016	3.3 x10 ⁻¹⁵	-0.124	0.016	3.7 x10 ⁻¹⁵	-0.054	0.016	7.9 x10 ⁻⁴	-0.054	0.016	8.1 x10 ⁻⁴
medium	-0.042	0.018	0.018	-0.042	0.018	0.018	-0.026	0.017	0.131	-0.026	0.017	0.13
high	0			0			0			0		

WGS * Educational attainment						
WGS * low	-0.037	0.01 6	0.018		-0.036	0.01 5 0.017
WGS * medium	-0.011	0.01 8	0.537		-0.009	0.01 7 0.588
WGS * high	0					

Abbreviations: eGFR: estimated glomerular filtration rate, WGS: weighted genetic score

^a Models with additional covariate adjustment (*models 5-10*) are presented in Web Table 1.

^b For longitudinal analysis, baseline eGFR was added to each model. Model 1: WGS + EA; Model 2: model 1 + WGS x EA; Model 3: WGS + EA + age + age² + sex + genetic principal components 1-10; Model 4: model 3 + WGS x EA

ORIGINAL UNEDITED MANUSCRIPT

TABLE 3. Comparison of nested ordinary least squares regression models with and without an interaction term for WGS x EA in PREVEND (1997-2012)

Model	R ²	R ² _{adj}	Res.Df	RSS	Δdf	F	Pr(>F) ^a
eGFR							
Model 1 ^b	0.064	0.063	3362	908742			
Model 2 ^c	0.064	0.063	3360	908379	2	0.670	0.512
Model 3 ^d	0.448	0.446	3349	535519			
Model 4 ^e	0.448	0.445	3347	535307	2	0.634	0.515
Model 5 ^f	0.463	0.459	3230	498513			
Model 6 ^g	0.463	0.459	3228	498367	2	0.475	0.622
Model 7 ^h	0.462	0.458	3195	485857			
Model 8 ⁱ	0.462	0.458	3193	485720	2	0.448	0.634
Model 9 ^{j,l}	0.462	0.458	3086	471547			
Model 10 ^{k,m}	0.462	0.458	3084	471432	2	0.377	0.686
annual eGFR change^c							
Model 1 ^b	0.041	0.040	3342	473.93			
Model 2 ^c	0.043	0.041	3340	473.03	2	3.177	0.042
Model 3 ^d	0.112	0.108	3329	438.62			
Model 4 ^e	0.114	0.109	3327	437.74	2	3.319	0.036
Model 5 ^f	0.130	0.124	3213	411.88			
Model 6 ^g	0.132	0.125	3211	411.17	2	2.777	0.062
Model 7 ^h	0.132	0.126	3178	405.28			
Model 8 ⁱ	0.133	0.126	3176	404.67	2	2.407	0.090
Model 9 ^{j,l}	0.132	0.125	3067	403.74			
Model 10 ^{k,m}	0.134	0.126	3065	402.94	2	3.026	0.049

Abbreviations: R², model explained variance; Res.Df, residual degrees of freedom; RSS, residual sum of squares; df, degrees of freedom; F, F-statistic.

^a P values Pr(>F) derived from F test using ANOVA between two nested models.

^bModel 1: (WGS + EA)

^cModel 2: model 1 + (WGS x EA)

^dModel 3: WGS + EA + age + age² + sex + genetic principal components (PC) 1-10

^eModel 4: model 3 + WGS x EA

^fModel 5: WGS + EA + age + age² + sex + PC 1-10 + BMI + SBP + glucose + total cholesterol + smoking

^gModel 6: model 5 + WGS x EA

^hModel 7: WGS + EA + age + age² + sex + PC 1-10 + BMI + SBP + glucose + total cholesterol + smoking + lnUAE

ⁱModel 8: model 7 + WGS x EA

^jModel 9: WGS + EA + age + age² + sex + PC 1-10 + BMI + hypertension + diabetes + total cholesterol + smoking + lnUAE

^kModel 10: model 9 + WGS x EA

¹ Diabetes (fasting glucose >7 mmol/L or non-fasting glucose >11 mmol/L or pharmacy-reported antidiabetic medication or self-reported diabetes) and hypertension (SBP >140 or DBP >90 or pharmacy-reported antihypertensive medication or self-reported hypertension)

^m For longitudinal analysis, baseline eGFR was included in each model.

ORIGINAL UNEDITED MANUSCRIPT

FIGURE 1. Plots of eGFR versus a Weighted Genetic Score for reduced eGFR, by educational attainment in PREVEND (1997-2012)

Upper panels show plots of cross-sectional estimated glomerular filtration rate (eGFR, mL/min/1.73m²) versus a Weighted Genetic Score, stratified by levels of educational attainment: high (panel **A**), medium (panel **B**), and low (panel **C**). Lower panels show plots of annual change in eGFR (mL/min/1.73m² per year) versus a Weighted Genetic Score, stratified by levels of educational attainment: high (panel **D**), medium (panel **E**), and low (panel **F**). Regression lines with 95% confidence interval are derived from unadjusted ordinary linear regression.

FIGURE 2. Multivariable adjusted associations of the Weighted Genetic Score with eGFR in strata of educational attainment in PREVEND (1997-2012)

Estimates of the associations, presented as regression coefficients with 95% confidence interval, of the Weighted Genetic Score (per standard deviation) with cross-sectional estimated glomerular filtration rate (eGFR, mL/min/1.73m²) (panel **A**) and annual eGFR change (mL/min/1.73m² per year) (panel **B**), derived from ordinary least squares regression analysis in the entire study population and in strata of educational attainment (high, medium, low). The solid lines represent an estimate of the interaction effect for the unadjusted model (model 11), while the dashed lines represent the interaction effect in the fully adjusted model (model 15, dashed) if it were linear.

Model 11^a: Weighted Genetic Score

Model 12^a: model 11 + age + age² + sex + genetic principal components 1-10

Model 13^a: model 12 + BMI + SBP + glucose + total cholesterol + smoking

Model 14^a: model 13 + lnUAE

Model 15^a: model 11 + BMI + hypertension + diabetes + total cholesterol + smoking + lnUAE

^afor longitudinal analysis, baseline eGFR was included in each model



