Chronic kidney disease

Thio, C. H. L.

DOI:
10.33612/diss.133648108

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.

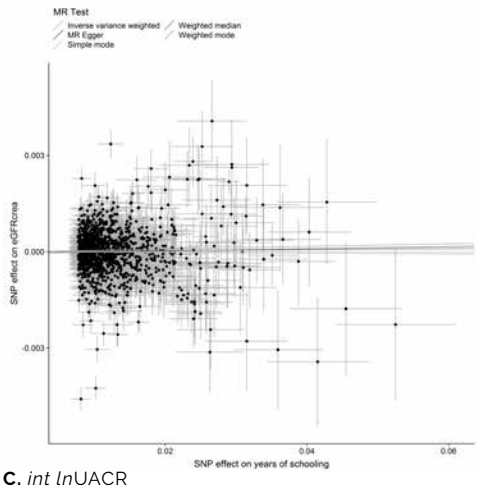
also through other traits, thereby violating MR assumptions and biasing the estimates). This was corroborated by sensitivity analyses that are robust to varying degrees of violations of MR assumptions (i.e. MR Egger, median and mode based methods, **Table 2**), as these yielded similar results. Additional 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

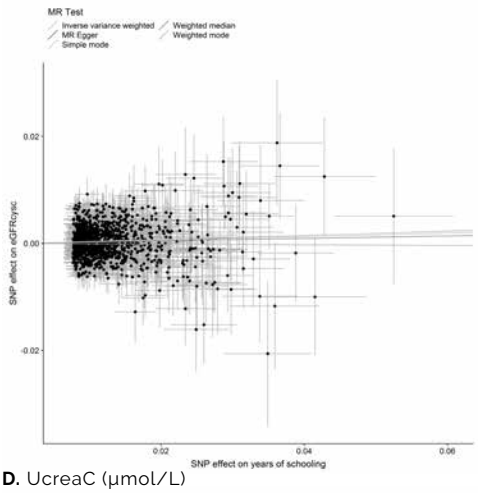
An effect of gEA on UACR was found, but this effect was not in the expected direction; each additional standard deviation in years of schooling (4.2 years) associated with an 0.06 increase in *int ln*UACR (IVW MR estimate: B=0.060, 95%CI: 0.043 to 0.077, p=2.5 $x10^{-12}$) (**Table 2, Figure 1**). Significant 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

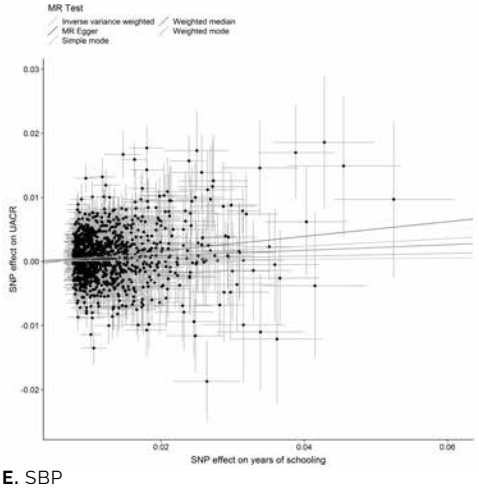
A. *lne*GFRcrea



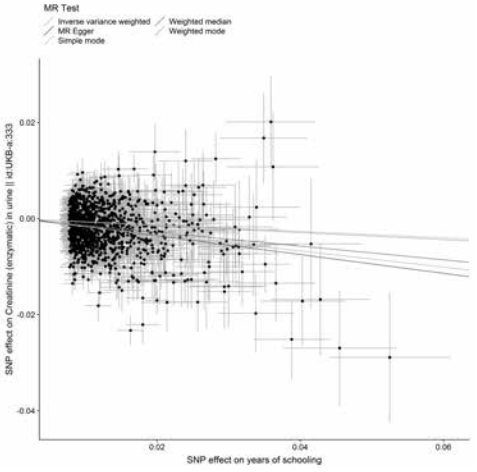
B. *lne*GFRcysc



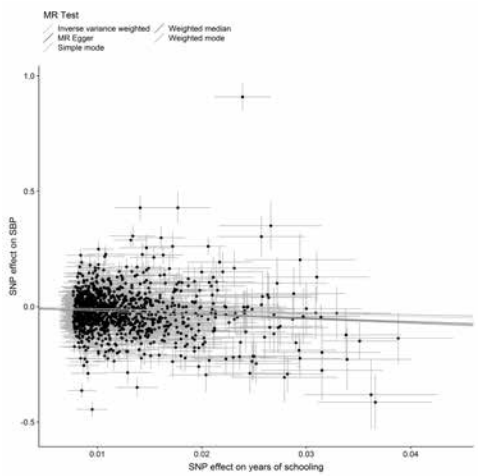
C. *int ln*UACR



D. UcreaC (µmol/L)

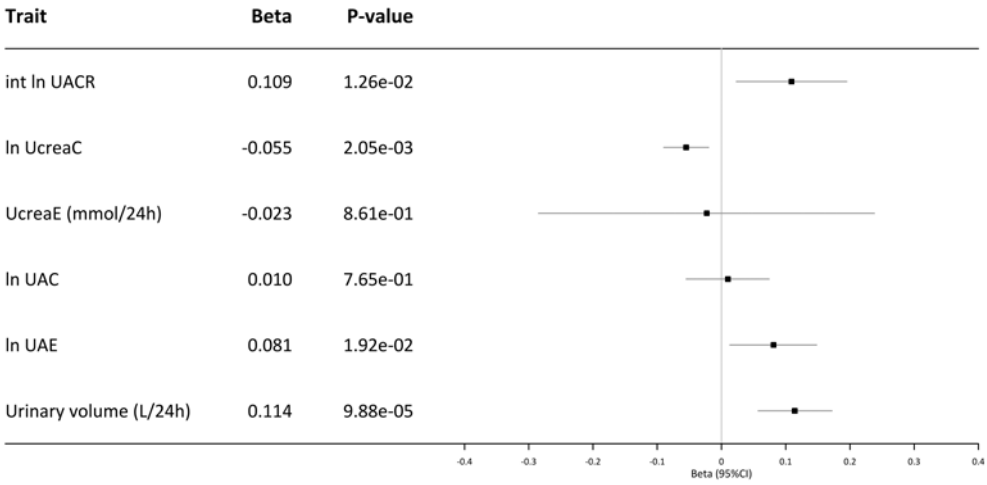


E. SBP



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.