Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

J Epidemiol Community Health

Supplementary Material

Roles of allostatic load, lifestyle and clinical risk factors in mediating the association between education and coronary heart disease risk in Europe.

Blánaid Hicks[*], Giovanni Veronesi, Marco M Ferrario, Hannah Forrest, Margaret Whitehead, Finn Diderichsen, Hugh Tunstall-Pedoe, Kari Kuulasmaa, Susana Sans, Veikko Salomaa, Barbara Thorand, Annette Peters, Stefan Soderberg, Giancarlo Cesana, Martin Bobak, Licia Iacoviello, Luigi Palmieri, Tanja Zeller, Stefan Blankenberg, Frank Kee; *On behalf of the MORGAM/BiomarCaRE consortium*

[*] Centre for Public Health, Queen's University Belfast. B.Hicks@qub.ac.uk

1

Supplementary Methods

The 3-way decomposition model and counterfactual definitions for its components

Let Y be a binary outcome, A a binary exposure (low education, A=a, and the reference high education, A=a*) and M a mediator (the allostatic load score, for instance). Let Ma and Ma* be the values assumed by the mediator in the two exposure levels. We present here the simplest situation, as an extensive body of literature has generalized these quantities to non-binary outcomes, exposures and mediators. At the individual counterfactual level, the total effect of the exposure on the outcome is $Y_a - Y_{a*}$, where Y_{a*} is the counterfactual outcome observed if we were able to change the exposure to the reference level. In the mediation analysis literature, the total effect is decomposed into the sum of n components, each representing a different mechanistic alternative through which the exposure affects the outcome. We were interested in decompositions that take into account the interaction between the exposure and the mediator, i.e. the 3-way [1] and the 4-way [2] decompositions. However, we deemed the former more appropriate to our case, since we expect the exposure to affect the levels of the mediator (i.e. education and allostatic load score are not independent).

The 3-way decomposition decomposes the total effect of the exposure on the outcome as the sum of three components [1]:

$$Y_a - Y_{a*} = (Y_{aMa*} - Y_{a*Ma*}) + (Y_{a*Ma} - Y_{a*Ma*}) +$$
$$(Y_{aMa} - Y_{aMa*} - Y_{a*Ma} + Y_{a*Ma*})(M_a - M_{a*})$$

The first term of the decomposition is the Pure Direct Effect, i.e. it is the effect of the exposure to low education assuming the mediator distribution remains the same as in the unexposed (i.e. the high education class).

The second term is the Pure Indirect Effect, i.e. it is the effect of changing the distribution of the mediator among the un-exposed to the levels of the exposed. This represents the effect of the exposure on the outcome that is entirely mediated by the differential exposure to the mediator.

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

J Epidemiol Community Health

The last term named Mediated Interaction is the product of an additive interaction between the exposure and the mediator on the outcome $(Y_{aMa} - Y_{aMa*} - Y_{a*Ma} + Y_{a*Ma*})$ and the effect of the exposure on the mediator $(M_a - M_{a*})$.

Use of marginal structural models with additive hazards regression to estimate the quantities of interest in the 3-way decomposition

Marginal structural models (MSM) are models for the marginal expectation of a counterfactual outcome [3]. Lange et al. [4] and Nordahl et al. [5] applied the MSM on the nested counterfactual outcome $Y_{a,Ma*}$ to estimate the quantities of interest for the 3-way decomposition. Using the same notation as above, and again at an individual counterfactual level, $Y_{a,Ma*}$ represents the nested counterfactual outcome for the exposure level A=a if we were able to change the distribution of the mediator M to that observed for the reference level A=a*. The general form for an MSM for $Y_{a,Ma*}$ is

$g(E[Y_{a,Ma*}]) = c_0 + c_1 a + c_2 a^* + c_3 a * a^*$, where g is a link function specifying the regression model for the outcome [4]. In the context of additive hazards models, it is required that the hazard function corresponding to the counterfactual survival time $Y_{a,Ma*}$ can be expressed as $\gamma_0(t) + c_1 a + c_2 a^* + c_3 a * a^*$, where $\gamma_0(t)$ is an unspecified baseline hazard [6]. Here the exposure is present three times to ascertain that it works through three distinct causal pathways, i.e. a direct effect (a), an indirect effect (a*), and their interaction. Conversely, there are no confounders as this is a marginal model [3]. The theoretical justification of the method and the counterfactual interpretation of c_1, c_2 and c_3 have been widely described [4]; the coefficients represent the average Pure Direct Effect, the average Pure Indirect Effect, and the average Mediated Interaction, respectively, and their sum is the total effect of the exposure on the outcome [5].

We report here the steps to estimate the coefficients $\hat{c}_1$, $\hat{c}_2$ and $\hat{c}_3$ in our paper. Firstly, we fitted linear (for allostatic load) or multinomial logistic (for smoking, alcohol intake and BMI) regression models for each mediator, given the educational class as the exposure, and adjusting for age and centre as potential confounders (C, in the density function below). All the models were gender-specific; results are in **Supplementary Tables S3** (men) **and S4** (women). From these models, one can estimate for the i-th individual the probability distribution for the mediator given the observed exposure and the covariates: $P(M = m \mid A = a_i, C = c_i)$; and the *counterfactual* probability distribution for the mediator if the exposure for the i-th individual was set to the level a*: $P(M = m \mid A = a_i^*, C = c_i)$. The counterfactual probability can be estimated by duplicating the available dataset as many times as the levels of the exposure (i.e. three times in our case), and creating an ancillary exposure A=a* for everyone. An appropriate subject's ID was created to allow estimation of robust standard errors in the additive hazards model. Then, for the i-th individual, weights are obtained as:

$$w_i = \frac{P(M = m_i \mid A = a_i^*, C = c_i)}{P(M = m_i \mid A = a_i, C = c_i)}$$

For a continuous mediator (i.e. allostatic load), the probabilities are replaced by the density functions, assuming a normal distribution. Weights are presented as **Supplementary Figures S1-S4**, to show that there are no "extreme" values (too small or too large weights) that might have produced instability in the final model (and to check the positivity assumption) [3]. Finally, the coefficients $\hat{c}_1$, $\hat{c}_2$ and $\hat{c}_3$ for the additive hazards model are estimated from weighted regression on the duplicated dataset, where a is the observed exposure level for everyone, and a* is the ancillary variable, using age as the underlying time scale, and adding the subject ID to a repeated statement to get robust standard error estimates. The Total Effect

4

is estimated as the sum of the three coefficients, and 95% confidence intervals were derived from robust standard error estimates and bootstrapping (n=100,000) [4-6]. We used the R (aalen function in the *timereg* package) and SAS programming statements in references [4-5] as examples for appropriate creation of the ancillary exposure, dataset duplication and weights estimation.

Interpretation of the estimated coefficients with the marginal structural model

The interpretation follows the work by Nordahl [5]. The average Pure Indirect Effect $\hat{c}_2$ rapresents *differential exposure*. The average Mediated Interaction $\hat{c}_3$ is the product of an additive interaction between the exposure and the mediator on the outcome, and the average effect of the exposure on the mediator, then $\hat{c}_3$ indicates *differential vulnerability*. The sum of $\hat{c}_2$ and $\hat{c}_3$ estimates the proportion eliminated by removing the mediator.

Goodness of fit for the additive hazards models

The additive hazards model $\gamma_0(t) + c_1 a + c_2 a^* + c_3 a * a^*$ assumes a constant effect of education on a common, "average" baseline hazard $\gamma_0(t)$. We found no evidence of a time-varying effect of education by using formal time-invariant test statistics in the *timereg* package [7]. For coronary heart disease, the baseline hazard $\gamma_0(t)$ increases strongly with age, and it is quite heterogeneous across the involved populations. Given that the aalen function in *timereg* estimates cumulative regression coefficients, the Total Effect of education estimated in Table 2 of the main text corresponds to a cumulative rate difference (events per 100,000 person years) between low and high education, in the 35-85 years interval of attained age during follow-up. These figures can be compared using the cumulative event rates, according to attained age during follow-up, and are reported in Table S5, in the overall sample and across different populations. In addition, we report in Table S6 the decomposition for allostatic load as mediator, after stratification according to 35-60 and 60-85 years of

attained age during follow-up. The Total Effect of low and intermediate education increases between the two age groups (e.g. men, low vs high education, 202 and 319 in 35-60 and 60-85 years old, respectively), although with mostly overlapping confidence intervals. Of note, the proportion of Mediated Interaction is larger in the youngest age group. Overall, the additive hazards model seems a reasonable approximation for our data.

Sensitivity analyses

To investigate the conditional exchangeability assumption, we performed sensitivity analyses to assess how confounding may have affected our *DE* (pure indirect effect) and *DS* (mediated interaction) estimates, based on simulations. We simulated an unmeasured continuous confounder U associated with the exposure (educational class) and the mediator (allostatic load [AL]). Simulations were carried out in both men and women, under different scenarios based on the combination of values for the U-E correlation (-0.2 to 0.0 by 0.05 increments) and U-M (AL) correlation (from -0.4 to 0.4 by 0.2 increments). Scenarios with positive U-E correlations yielded the same results and are not reported here. The negative U-E correlation means that the distribution of U increases in less educated individuals, and vice-versa. We added U to the models for estimating the weights for the marginal structural models as described above, to estimate DE and DS for AL. The average results over 40 simulation runs are displayed in Supplementary **Figure S5** (males) and **Figure S6** (females).

In the figures, the point labelled with "A" [U-E correlation and U-AL correlation = 0] corresponds to the observed estimate for DE (DS) in our study (as presented in Table 2). The remaining points represent different simulation scenarios (for U-E and U-AL correlation values). The point with the label "C" corresponds to a confounder U with a moderate negative correlation with E, and a weak positive correlation with AL. Based on the paper by

6

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

J Epidemiol Community Health

Ribeiro et al [8], such a confounder broadly corresponds to neighbourhood socio-economic deprivation (less educated subjects are more likely to live in most deprived neighbourhood; and those living in most deprived neighbourhoods have higher AL). Thus, moving from "A" to "C" gives a sense of the amount of bias in our data due to having neglected a confounder with the same characteristics of neighbourhood deprivation.

The point with the label "B" corresponds to a confounder U with a weak negative correlation with E, and a moderate to strong positive correlation with AL. In our data, such a confounder would have the characteristics of age. Thus, moving from "A" to "B" gives a sense of the amount of bias in our data due to having neglected a confounder with the same characteristics of age.

Online method references

1. VanderWeele TJ. A three-way decomposition of a total effect into direct, indirect, and interactive effects. Epidemiology. 2013;24(2):224–232. doi:10.1097/EDE.0b013e318281a64e

2. VanderWeele TJ. A unification of mediation and interaction: a 4-way decomposition [published correction appears in Epidemiology. 2016 Sep;27(5):e36]. Epidemiology. 2014;25(5):749–761. doi:10.1097/EDE.0000000000000121

3. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000;11(5):550–560. doi:10.1097/00001648-200009000-00011

4. Lange T, Vansteelandt S, Bekaert M. A simple unified approach for estimating natural direct and indirect effects. Am J Epidemiol. 2012;176(3):190–195. doi:10.1093/aje/kwr525

5. Nordahl H, Lange T, Osler M, et al. Education and cause-specific mortality: the mediating role of differential exposure and vulnerability to behavioral risk factors. Epidemiology. 2014;25(3):389–396. doi:10.1097/EDE.0000000000000080

6. Lange T, Hansen JV. Direct and indirect effects in a survival context. Epidemiology. 2011;22(4):575–581. doi:10.1097/EDE.0b013e31821c680c

7. Martinussen T, Scheike TH. Dynamic regression models for survival data. New York, NY: Springer; 2006.

8. Ribeiro AI, Fraga S, Kelly-Irving M, et al. Neighbourhood socioeconomic deprivation and allostatic load: a multi-cohort study. *Sci Rep*. 2019;9(1):8790.

8

Table S1: Characteristics of the surveyed populations, number of subjects, and percentage of subjects with complete data by educational class

Region	Population	No. of cohorts	Setting	Baseline period	Part rates	No. of subjects†		Follow-up length and events no.	
						M	W	Years°°	CHD
Nordic Countries	Northern Sweden (Västerbotten\Norrbotten Counties)°	6	U/R	1999-09	75%	3172	3476	12.7	394
	East Finland-FINRISK (North Karelia\Kuopio)	1	R	1997	75%	909	1018	13.8	135
	West Finland-FINRISK (Helsinki\Turku\Loimaa)	1	U	1997	71%	812	873	13.8	90
The UK	Northern Ireland (PRIME-Belfast#)	1	U	1991-94	52%	1977	-	18.0	228
	Scottish Heart Health Extended Cohorts (SHHEC)^	4	U/R	1984-95	70%	5799	6059	22.7	1637
Central and South Europe	Germany (MONICA/KORA Augsburg)	1	U/R	1994-95	74%	1551	1523	14.0	193
	Northern Italy (Brianza)	3	U	1986-94	67%	1612	1717	18.9	194
	Southern Italy (Latina)‡	2	R	1993-96	56%	241	847	10.5	24
	Southern Italy (Moli-Sani)	1	U/R	2005-10	70%	8390	9574	4.3	98
	Spain (Catalonia)	1	U/R	1986-88	74%	880	931	9.8	38
	All populations	**21**	-	-	-	**25310**	**26018**	**10.1**	**3031**

Setting: U = Urban, R = Rural. Part rates: participation rates, computed from responders and invited in every survey of a given population. In case of re-examinations, participation is referred to the initial survey.

°: 3 surveys with baseline visits in 1999, 2004 and 2009; and the 1999 re-examination of 3 additional surveys with original baseline visit in 1986, 1990 and 1994.

#: survey included into the PRIME study. The survey enrolled only men aged 49-60 years at baseline.

^: MONICA Glasgow, MONICA Edinburgh and Scottish Heart Health Study. ‡: re-examination of the original surveys recruited in 1983-87.

†: 35-74 years old with data on education, allostatic load and on the covariates of interest, and free of CVD at baseline.

°°: median follow time

9

Table S2: Measurement details for the markers involved into the allostatic load score definition.

AL score marker	Where it was measured	Unit	Material	Measurement details
C-Reactive Protein	Local laboratory (lab) for Augsburg; centralized lab for the remaining populations	mg/L	Serum	-
Glucose	Local lab for Belfast; centralized lab for the remaining populations	mmol/L	Serum	Fasting status: Non-fasting specimens in the SHHEC study. Mixture of overnight fasting, 4-hour fasting and non-fasting samples in FINRISK study. Northern Sweden: Approximately 60% of all participants had an overnight fast, the remaining 40% at least 4 hours fasting. Overnight fasting observed in the remaining populations.
HbA1C	Local lab (available only for Augsburg, Brianza and Northern Sweden)	mmol/mol	Whole blood	Relevant for the KORA-Augsburg cohort only.
TC	Local lab for all the populations	mmol/L	Serum/Plasma[1]	DQA available at: https://www.thl.fi/morgam/a/publications/qa/baseline/chol/cholqa.htm
HDL-C	Centralized lab for Northern Sweden; local lab for the remaining populations	mmol/L	Serum/Plasma[1]	DQA available at: https://www.thl.fi/morgam/a/publications/qa/baseline/chol/cholqa.htm
Triglycerides	Centralized lab for Northern Sweden; local lab for the remaining populations	mmol/L	Serum/Plasma[1]	DQA available at: https://www.thl.fi/morgam/a/publications/qa/baseline/chol/cholqa.htm. Fasting status: Non-fasting specimens in the SHHEC study. Mainly non-fasting in MONICA/KORA Augsburg. Mixture of overnight fasting, 4-hour fasting and non-fasting samples in FINRISK study. Northern Sweden: Approximately 60% of all participants had an overnight fast, the remaining 40% at least 4 hours fasting. Overnight fasting observed in the remaining populations.
Blood pressure (systolic and diastolic)	Local measurement	mmHg	NA	Blood pressure was measured after 2-5 minutes rest while sitting, using a standard or random zero mercury sphygmomanometer or an automated oscillometric device. With the exception of Belfast (one measure only), two consecutive measurements were available, and the average was used as the study variable for systolic and diastolic blood pressure.
Body Mass Index	Local measurement of height and weight	Kg/m2	NA	We computed the Body Mass Index Quetelet index from measured height and weight.

10

Local: measurements performed locally by each population
Centralized: measurements performed in the MORGAM/BiomarCaRE consortium laboratory. The laboratory was firstly located at the Johannes Gutenberg University, Mainz, and then moved at the University Heart Center, Hamburg. DQA = Data Quality Assessment
[1]: plasma measure only for the Northern Ireland cohort

11

Table S3: Fitted linear or multinomial logistic regression models for the different mediators (allostatic load score, smoking, alcohol and BMI) conditioning on exposure (education), age and study center among 25310 men, 35-74 years old and CVD-free at baseline

	Low vs. high education			Intermediate vs. high education		
	Estimate	**Std. Error**	**p-value**	**Estimate**	**Std. Error**	**p-value**
Allostatic Load	0.5978	0.0586	<.0001	0.4981	0.0651	<.0001
Smoking						
Never smokers	-0.2193	0.0352	<.0001	-0.2376	0.0387	<.0001
Former smokers	Ref			Ref		
1-10 cigs/day	0.1460	0.0626	0.0198	0.1243	0.0687	0.0706
11-20 cigs/day	0.5049	0.0475	<.0001	0.2762	0.0536	<.0001
>20 cigs/day	0.5915	0.0651	<.0001	0.3944	0.0731	<.0001
Alcohol intake						
0 (Teetotallers)	Ref	-	-	Ref	-	-
1-2 drinks/day	-0.3808	0.0407	<.0001	-0.1558	0.0456	0.0006
3-4 drinks/day	0.1096	0.0451	0.0150	0.1816	0.0504	0.0003
5 or more	0.9416	0.0839	<.0001	0.7915	0.0933	<.0001
Body Mass Index						
Normal weight	Ref			Ref		
Overweight	0.1404	0.0346	<.0001	0.1364	0.0384	0.0004
Obese	0.5249	0.0441	<.0001	0.4184	0.0491	<.0001

In the table: beta-coefficient (estimate), standard error and p-value (Wald chi-square test) for testing the null hypothesis of no association between educational class and the mediator value. Models are adjusted for age and study center. Ref=reference category (multinomial logistic regression models)

12

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

J Epidemiol Community Health

Table S4: Fitted linear or multinomial logistic regression models for the different mediators (allostatic load score, smoking, alcohol and BMI) conditioning on exposure (education), age and study center among 26018 women, 35-74 years old and CVD-free at baseline

	Low vs. high education			Intermediate vs. high education		
	Estimate	**Std. Error**	**p-value**	**Estimate**	**Std. Error**	**p-value**
Allostatic Load	1.3470	0.0577	<.0001	0.7675	0.0636	<.0001
Smoking						
Never smokers	0.5098	0.0408	<.0001	0.2746	0.0432	<.0001
Former smokers	Ref	-	-	Ref	-	-
1-10 cigs/day	0.5005	0.0585	<.0001	0.2232	0.0639	0.0005
More than 11 cigs/day	0.8692	0.0555	<.0001	0.4260	0.0634	<.0001
Alcohol intake						
0 (Teetotallers)	Ref			Ref		
1-2 drinks/day	-0.4853	0.0306	<.0001	-0.2033	0.0336	<.0001
More than 3 drinks/day	-0.2716	0.0683	<.0001	-0.1381	0.0749	0.0651
Body Mass Index						
Normal weight	Ref			Ref		
Overweight	0.4975	0.0347	<.0001	0.2929	0.0377	<.0001
Obese	0.9320	0.0402	<.0001	0.5296	0.0438	<.0001

In the table: beta-coefficient (estimate), standard error and p-value (Wald chi-square test) for testing the null hypothesis of no association between educational class and the mediator value. Models are adjusted for age and study center. Ref=reference category (multinomial logistic regression models)

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

J Epidemiol Community Health

Table S5: Estimated cumulative coronary heart disease event rates (per 100,000 person-years) at selected attained ages during follow-up, in the included populations and overall estimate. Men (left) and women (right).

Population	Men					Women				
	Attained age during follow-up					Attained age during follow-up				
	40	50	60	70	80	40	50	60	70	80
N Sweden	183	368	740	1491	3002	50	115	267	619	1434
Finland	127	271	575	1222	2598	35	88	221	557	1403
UK-Belfast	na	692	759	832	912	na	na	na	na	na
Scotland	265	476	855	1536	2760	65	150	349	812	1888
Augsburg	78	185	441	1050	2499	31	81	213	559	1465
Brianza	203	321	508	803	1270	24	59	142	344	832
Latina	61	133	291	637	1395	4	16	62	235	891
Moli-sani	75	134	239	427	761	10	23	55	132	317
Catalonia	104	210	424	859	1222^	5	22	104	483	1042^
All populations	186	352	666	1259	2381	40	100	248	618	1537

In the table: age-adjusted cumulative event rates in the populations, estimated from Poisson models adjusting for attained age during follow-up.
^: estimated at the attained age during follow-up of 75 (latest observed value: 77 in men and 76 in women). In women we observed only 1 event after the attained age of 70.
na: not available (no participants)

14

Table S6: Rate Difference in additional coronary heart disease events per 100,000 person-years by educational Level (decomposition of TE into direct, Indirect and Mediated Interaction effects) for allostatic load score as mediator, stratified by attained ages during follow-up.

Mediator=Allostatic Load score[1]	Men			Women		
	Educational Level[2]		Proportion mediated (95%CI)	Educational Level[2]		Proportion mediated (95%CI)
	Low	Intermediate		Low	Intermediate	
	RD (95% CI)	RD (95% CI)		RD (95% CI)	RD (95% CI)	
Attained age 35-60						
Total Effect	202 (131 to 272)	83 (10 to 156)		102 (63 to 140)	26 (-11 to 63)	
Pure Direct Effect (PDE)	167 (98 to 235)	63 (-9 to 134)		64 (29 to 99)	14 (-24 to 52)	
Pure Indirect Effect (PIE)	25 (19 to 30)	25 (20 to 29)	12 (8 to 20)	22 (14 to 31)	16 (13 to 20)	22 (11 to 42)
Mediated Interaction (MI)	10 (3 to 18)	-4 (-9 to 0)	5 (2 to 9)	15 (4 to 27)	-5 (-8 to -1)	15 (5 to 24)
Attained age 60-85						
Total Effect	319 (186 to 452)	195 (64 to 326)		230 (139 to 322)	120 (19 to 221)	
Pure Direct Effect (PDE)	262 (129 to 394)	153 (23 to 283)		152 (61 to 242)	81 (-20 to 183)	
Pure Indirect Effect (PIE)	54 (45 to 64)	47 (40 to 55)	17 (11 to 31)	71 (53 to 89)	40 (32 to 48)	31 (19 to 56)
Mediated Interaction (MI)	3 (-9 to 15)	-5 (-13 to 3)	1 (-3 to 5)	8 (-15 to 31)	-1 (-9 to 6)	4 (-8 to 13)

1: Analyses adjusted for age and center
2: reference category: high education
RD: Risk difference, estimated from Additive Hazard survival model, with age on the time scale and adjusting for center
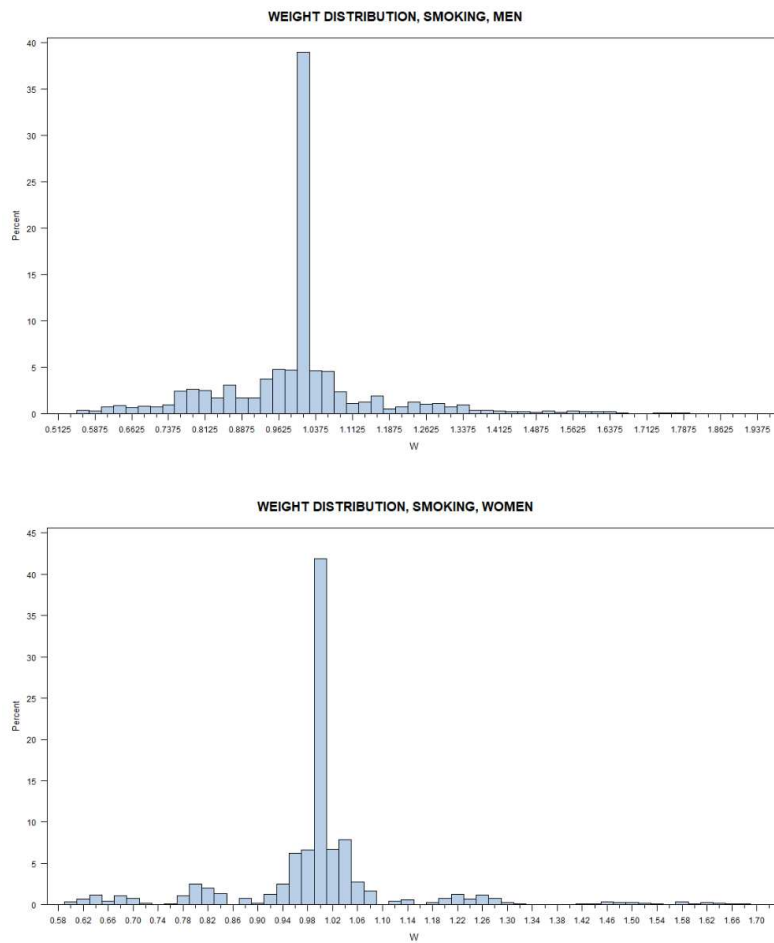
15

Figure S1: Histogram of IPTW weights for allostatic load, in men (above, n=25310) and women (below, n=26018).
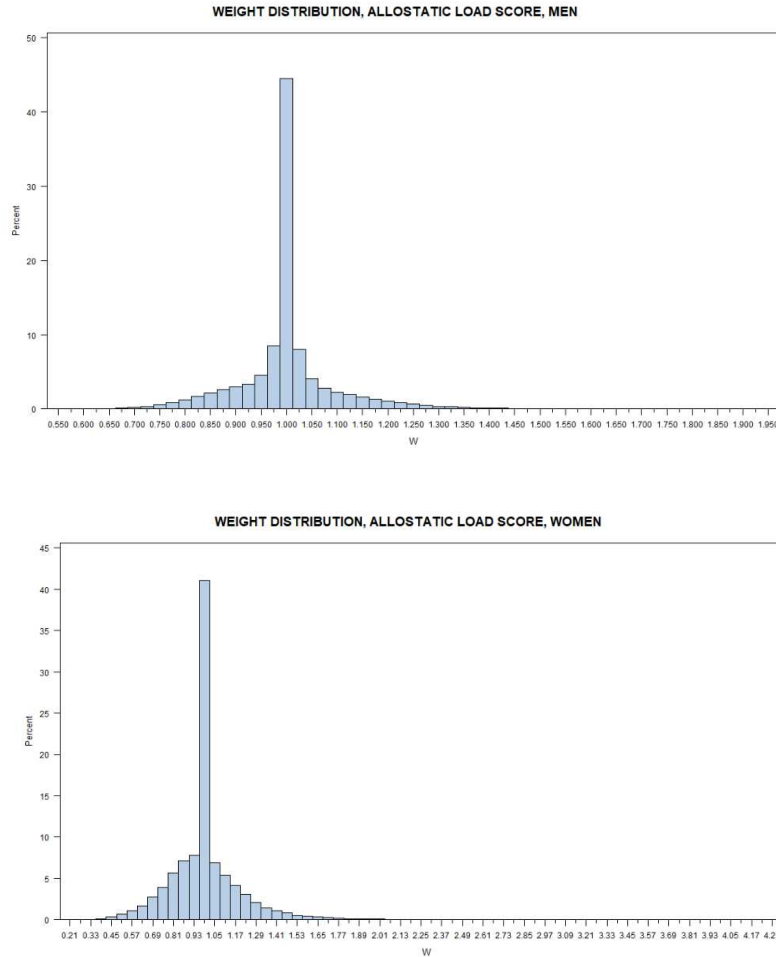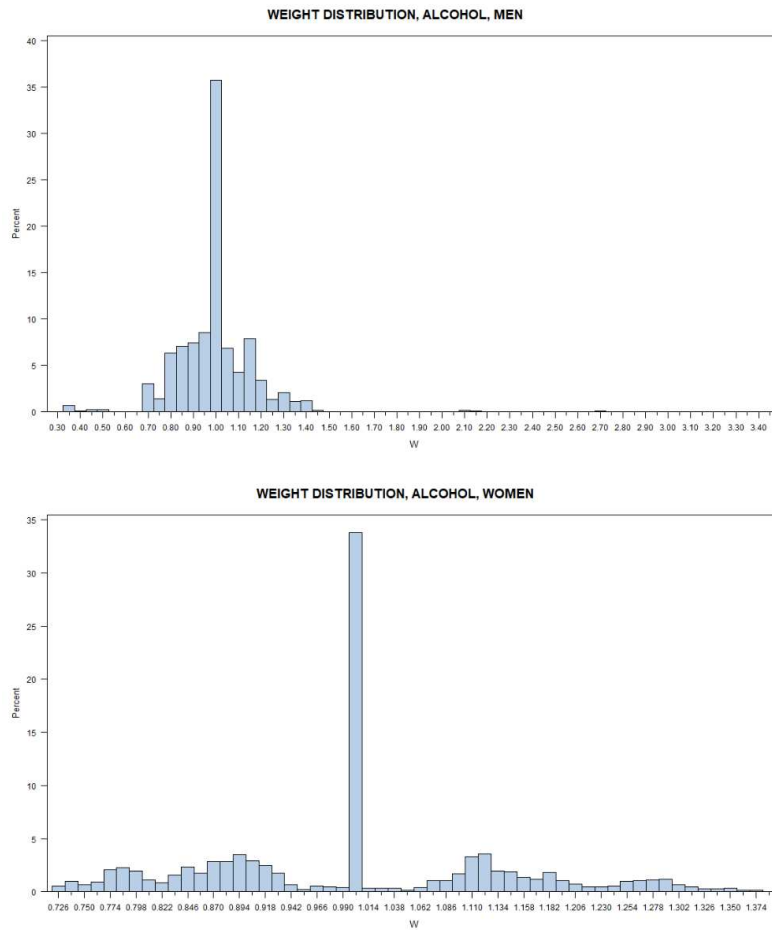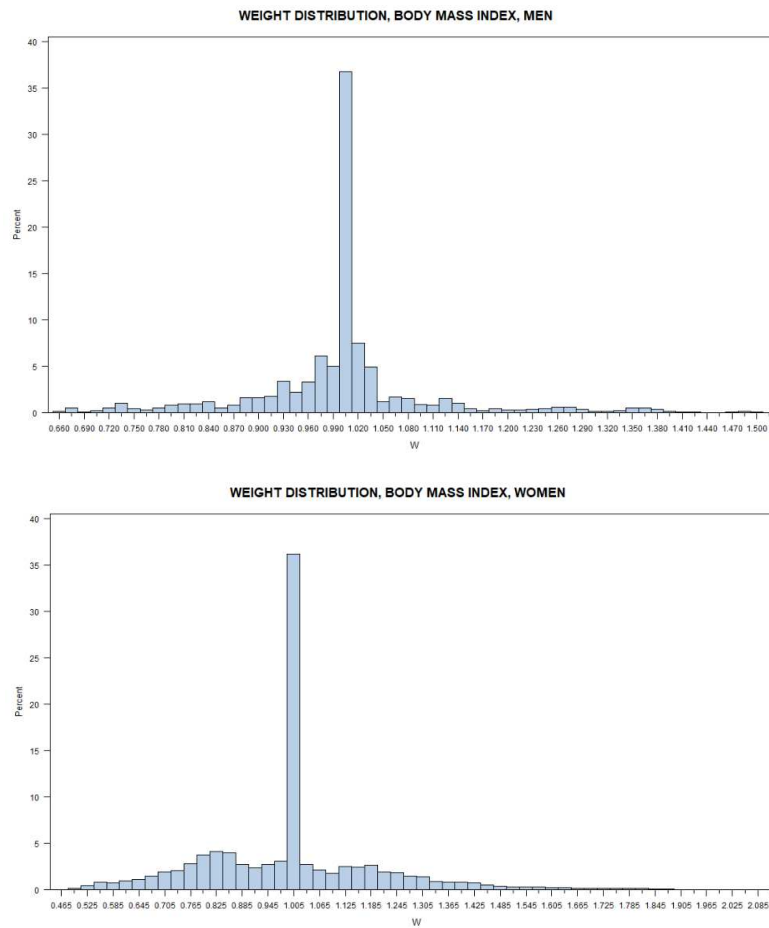

*IPTW; Inverse Probability of Treatment Weights

Figure S2: Histogram of weights for smoking, W, in men (above, n=25310) and women (below, n=26018).

Figure S3: Histogram of weight for alcohol intake, W, in men (above, n=25310) and women (below, n=26018).

Figure S4: Histogram of weight for body mass index, W, in men (above, n=25310) and women (below, n=26018).

WEIGHT DISTRIBUTION, BODY MASS INDEX, MEN

WEIGHT DISTRIBUTION, BODY MASS INDEX, WOMEN

19

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

J Epidemiol Community Health

Figure S5 Sensitivity analysis investigating the effects of unmeasured confounding on *differential exposure* and *susceptibility* estimates men
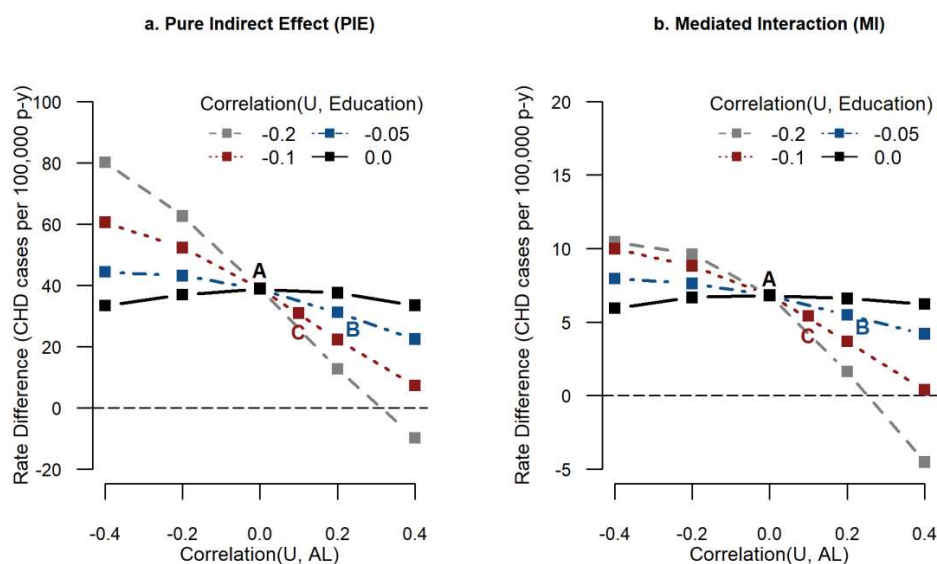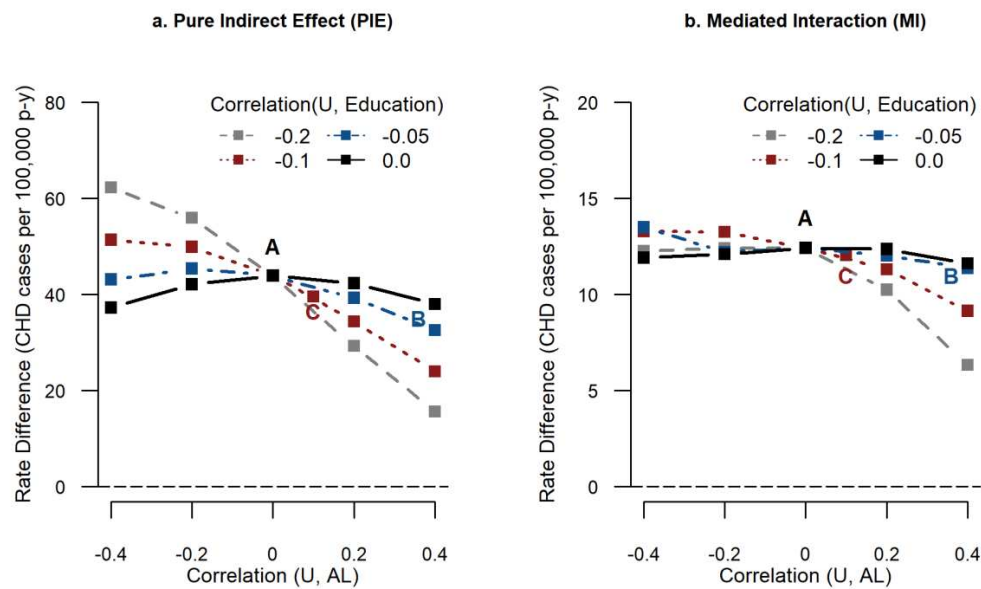

PIE= Pure indirect effect (*differential exposure*); MI= mediated interaction (*differential susceptibility*).
The points labeled with "A" is the observed estimates in our study for men (as presented in Table 2). The remaining points represent different simulation scenarios (for U-E and U-AL correlation values).
The point with the label "C" corresponds to a confounder U with a moderate negative correlation with Education, and a weak positive correlation with AL. Based on the paper by Ribeiro et al (5), such a confounder broadly corresponds to neighborhood socio-economic deprivation. Thus, moving from "A" to "C" gives a sense of the amount of bias in our data due to having neglected a confounder with the same characteristics of neighborhood deprivation.
The point with the label "B" corresponds to a confounder U with a weak negative correlation with E, and a moderate positive correlation with AL. In our data, such a confounder has the characteristics of age. Thus, moving from "A" to "B" gives a sense of the amount of bias in our data due to having neglected a confounder with the same characteristics of age.

20

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

J Epidemiol Community Health

Figure S6 Sensitivity analysis investigating the effects of unmeasured confounding on *differential exposure* and *susceptibility* estimates in women

PIE= Pure indirect effect (*differential exposure*); MI= mediated interaction (*differential susceptibility*).
The points labeled with "A" is the observed estimates in our study for women (as presented in Table 2). The remaining points represent different simulation scenarios (for U-E and U-AL correlation values).
The point with the label "C" corresponds to a confounder U with a moderate negative correlation with Education, and a weak positive correlation with AL. Based on the paper by Ribeiro et al (5), such a confounder broadly corresponds to neighborhood socio-economic deprivation. Thus, moving from "A" to "C" gives a sense of the amount of bias in our data due to having neglected a confounder with the same characteristics of neighborhood deprivation.
The point with the label "B" corresponds to a confounder U with a weak negative correlation with E, and a strong positive correlation with AL. In our data, such a confounder has the characteristics of age. Thus, moving from "A" to "B" gives a sense of the amount of bias in our data due to having neglected a confounder with the same characteristics of age.