

# SCIENTIFIC REPORTS



OPEN

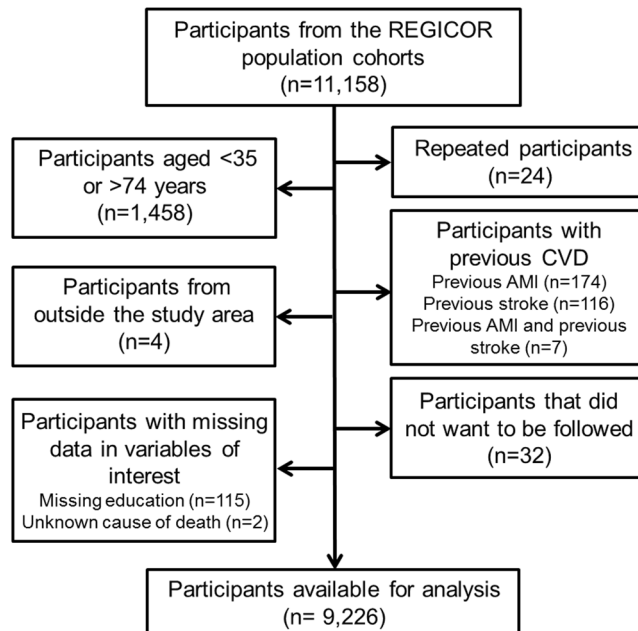
## The association between education and cardiovascular disease incidence is mediated by hypertension, diabetes, and body mass index

Irene R. Dégano<sup>1,2</sup>, Jaume Marrugat<sup>1,2</sup>, Maria Grau<sup>2,3</sup>, Betlem Salvador-González<sup>2,4</sup>, Rafel Ramos<sup>5,6,7</sup>, Alberto Zamora<sup>6,8</sup>, Ruth Martí<sup>5,9</sup> & Roberto Elosua<sup>1,2</sup>

Education and cardiovascular disease (CVD) are inversely associated but the mediating factors have not been totally elucidated. Our aim was to analyze the mediating role of modifiable risk factors. Cohort study using the REGICOR population cohorts. Participants without previous CVD were included ( $n = 9226$ ). Marginal structural models were used to analyze the association between education and CVD incidence at 6 years of follow-up. Mediation by modifiable risk factors (diabetes, dyslipidemia, hypertension, smoking, body mass index, and physical activity) was assessed using the counterfactual framework. Participants with a university degree had a CVD incidence hazard ratio (HR) of 0.51 (95% confidence interval (CI) = 0.30, 0.85), compared to those with primary or lower education. Only hypertension, BMI, and diabetes mediated the association between education and CVD incidence, accounting for 26% of the association (13.9, 6.9, and 5.2%, respectively). Sensitivity analyses showed that hypertension was the strongest mediator (average causal mediation effect [95% CI] = increase of 2170 days free of CVD events [711, 4520]). The association between education and CVD incidence is partially mediated by hypertension, BMI, and diabetes. Interventions to decrease the prevalence of these risk factors could contribute to diminish the CVD inequalities associated with educational level.

There are a number of determinants that affect health. The World Health Organization has defined that determinants of health include the social and economic environment, the physical environment, and the individual characteristics and behaviors<sup>1</sup>. The more specific drivers are 8 including education, income, and social status. Regarding cardiovascular disease (CVD) it has been observed that the factors that have been involved in the decrease in CVD mortality, such as prevention and treatment, have not equally reached all social groups<sup>2</sup>. And more importantly, the opportunities to reduce morbidity and mortality due to CVD are associated to addressing the social determinants of CVD as disparities have increased over time<sup>3</sup>. It is therefore essential to understand the association of social determinants not only with disease incidence but also with disease risk factors as these can be targeted with specific interventions. In the last decades an inverse association between socioeconomic status (SES) and CVD has been observed<sup>4</sup>, and a number of studies have shown that lower SES is associated with higher

<sup>1</sup>CIBER of Cardiovascular Diseases (CIBERCV), ISCIII, Madrid, Spain. <sup>2</sup>REGICOR Study Group. Cardiovascular Epidemiology and Genetics Group. Epidemiology and Public Health Program. IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain. <sup>3</sup>School of Pharmacy, University of Barcelona, Barcelona, Spain. <sup>4</sup>ABS Florida Sud and Cardiovascular Research Group in Primary Care (MACAP Costa Ponent), Primary Care Research Institute Jordi Gol, Catalan Institute of Health, l'Hospitalet de Llobregat, Barcelona, Spain. <sup>5</sup>Vascular Health Research Group (ISV-Girona). Institut Universitari d'Investigació en Atenció Primària Jordi Gol (IDIAP Jordi Gol), Barcelona, Spain. <sup>6</sup>Translab Research Group, Department of Medical Sciences, School of Medicine, University of Girona, Girona, Spain. <sup>7</sup>Institut Català de la Salut, Àmbit d'atenció Primària, Girona, Spain. <sup>8</sup>Lipid and Atherosclerosis Unit and Department of Internal Medicine, Hospital de Blanes, Girona, Spain. <sup>9</sup>Institut d'Investigació Biomèdica de Girona (IDIBGI), Hospital Universitari Dr. Josep Trueta, Girona, Spain. Correspondence and requests for materials should be addressed to I.R.D. (email: [iroman@imim.es](mailto:iroman@imim.es)) or R.E. (email: [relosua@imim.es](mailto:relosua@imim.es))



**Figure 1.** Flow chart of the participants included in the study.

risk of CVD<sup>5–10</sup>. However, which are the factors explaining this association is not entirely clear. Identifying the mediating factors is the first step to diminish CVD inequalities associated to different SES, particularly regarding modifiable factors, which can be tuned with prevention such as lifestyle interventions on diet, physical activity, and smoking, and pharmacological treatment.

A number of studies have analyzed the factors underlying the association between SES and CVD<sup>11–17</sup>. And more recently some authors have examined the mediating role of modifiable classical cardiovascular risk factors (hypertension, diabetes, and dyslipidemia) and/or lifestyle factors (smoking, physical activity -PA-, body mass index -BMI-, and diet)<sup>18–20</sup>. All these studies have shown evidence of the implication of both classical and lifestyle modifiable risk factors on the association between SES and CVD but their results were highly heterogeneous. In addition, these studies have not used new methodologies to estimate the effect of SES on CVD through a set of mediators and have not evaluated whether awareness, treatment, or control of classical risk factors is playing a role.

To fill this gap, the present study used a large population based cohort<sup>21,22</sup> to analyze the mediating role of modifiable risk factors on the association between education, as a SES indicator, and 6-year CVD incidence.

## Results

**Descriptive analysis by education level.** From the 9226 included participants (Fig. 1), 5322 (58%) received primary (4990) or lower education (332), 2323 (25%) reached secondary education, and 1581 (17%) obtained a university degree. More educated participants were significantly younger and more frequently men, had a lower prevalence of diabetes, hypertension and dyslipidemia, and a higher prevalence of smoking (Table 1). Higher education was also associated with lower BMI, more vigorous PA practice, higher perceived physical health and with having a non-manual occupation. The percentage of participants who had a CV event in the follow-up diminished from the less to the most educated (5.7–2.9–2.2%).

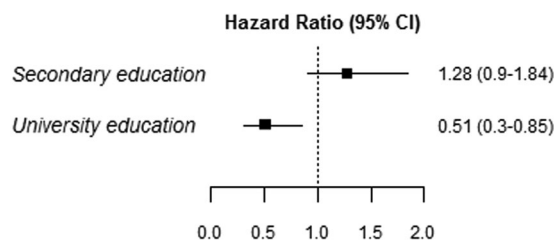
**Association between education and CVD incidence.** Participants who obtained a university degree had a CV event hazard ratio (HR) of 0.51 (95% confidence interval (CI): 0.30–0.85), compared to the ones who received primary or lower education (Fig. 2). No association was observed when comparing participants that completed secondary education only to the ones that received primary or lower education. There was no effect modification by gender, while there was a slight effect modification by age in participants who reached secondary education only (Suppl. Table S1, Suppl. Fig. S1).

Sensitivity analyses for exchangeability and unmeasured confounding by categorical variables showed that those were not a concern (Suppl. Results and Tables S2 and S3). Sensitivity analyses for positivity and correct model specification assumptions showed similar results (Suppl. Results and Tables S2 and S4).

**Mediation of the association between education and CVD incidence.** Hypertension, BMI, and diabetes were identified as mediators of the association between education and 6-year CVD incidence (Table 2). The average causal mediation effects (ACMEs) or indirect effects shown in Table 2 are the expected difference in the number of days without a CVD event when the mediator takes the value observed in individuals with university education compared to those with primary or lower education holding education level constant. Thus, an increase of 2170 and 815 days without a CVD event was expected when comparing hypertension and diabetes prevalence of the most educated to the prevalence in the less educated. Similarly, an increase of 1080 days without a CVD

	Elementary education	Secondary Education	University Education	p-trend	N
	N = 5,322	N = 2,323	N = 1,581		
Age (years) <sup>a</sup>	56.8 (10.8)	49.9 (9.9)	49.4 (9.7)	<0.001	9226
Sex (% female)	2,885 (54.2%)	1,241 (53.4%)	768 (48.6%)	<0.001	9226
Diabetes	853 (16.0%)	234 (10.1%)	129 (8.2%)	<0.001	9226
Dyslipidemia	2,426 (45.6%)	832 (35.8%)	523 (33.1%)	<0.001	9226
Hypertension	2,492 (46.8%)	677 (29.1%)	391 (24.7%)	<0.001	9226
Smoking				<0.001	9117
Non-smokers	3,138 (59.7%)	1,095 (47.5%)	700 (44.9%)		
Ex-smokers	1,053 (20.0%)	530 (23.0%)	445 (28.5%)		
Current smokers	1,063 (20.2%)	679 (29.5%)	414 (26.6%)		
BMI (kg/m <sup>2</sup> ) <sup>a</sup>	28.1 (4.6)	26.6 (4.4)	25.9 (4.2)	<0.001	9151
Light PA (METs) <sup>a</sup>	96.8 (131)	76.2 (105)	70.2 (95.8)	<0.001	9135
Moderate PA (METs) <sup>a</sup>	107 (206)	88.9 (158)	89.3 (123)	<0.001	9135
Vigorous PA (METs) <sup>a</sup>	92.6 (203)	126 (195)	147 (215)	<0.001	9135
Total PA (METs) <sup>a</sup>	297 (337)	291 (290)	306 (282)	0.339	9135
Inactive	4,895 (93.0%)	2,118 (92.1%)	1,444 (91.9%)	0.196	9135
Social class (% manual)	3,948 (79.7%)	948 (41.9%)	129 (8.3%)	<0.001	8774
Quality of life – PCS <sup>a</sup>	48.4 (9.3)	51.6 (7.8)	53.0 (6.9)	<0.001	8500
Quality of life – MCS <sup>a</sup>	47.6 (11.2)	47.3 (10.5)	47.6 (10.0)	0.650	8500
Cohort				<0.001	9226
1995	1,068 (20.1%)	204 (8.8%)	66 (4.2%)		
2000	1,740 (32.7%)	497 (21.4%)	222 (14.0%)		
2005	2,514 (47.2%)	1,622 (69.8%)	1,293 (81.8%)		
Follow-up (days) <sup>a</sup>	3,108 (1,178)	2,754 (991)	2,493 (749)	<0.001	9226
CVD events	304 (5.7%)	67 (2.9%)	34 (2.2%)	<0.001	9226

**Table 1.** Characteristics of the included participants by education level. N(%) is shown except for <sup>a</sup>were mean and standard deviation is presented. P-trends were obtained with the Pearson test. BMI: body mass index; CVD: cardiovascular disease MCS: mental component score from the SF-36 questionnaire; PA: physical activity; PCS: physical component score from the SF-36 questionnaire.



**Figure 2.** Hazard ratio (HR) and 95% confidence interval (CI) of cardiovascular disease incidence for participants with secondary and university education compared to participants with primary or lower education. The association between education and CVD incidence was analyzed with a marginal structural survival model. Inverse probability weights and robust standard errors were included in the model.

event was expected when comparing mean BMI of the most educated to the value in the less educated. The total effect of education on CVD incidence was 15600 days free of CVD events and the direct effects of the mediators analyzed are presented in Suppl. Table S5. The percentage of the education-CVD incidence association mediated by the identified mediators was 26%. Hypertension, BMI, and diabetes mediated 13.9%, 6.9%, and 5.2% of the association, respectively.

These results were consistent in several sensitivity analyses for residual confounding and measurement errors (Suppl. Table S5). Including education-mediators interaction terms in the outcome models pointed to hypertension as the strongest mediator (Suppl. Table S5).

Awareness, treatment, and control of diabetes and hypertension in individuals with these risk factors were not playing a role in mediating the association between education and 6-year CVD incidence (Suppl. Table S6).

## Discussion

Our study showed an inverse association between education and CVD incidence in general population from North-Eastern Spain when comparing participants with university education to participants with primary or lower education. Mediation analyses identified that 26% of this association was mediated by classical and lifestyle

Mediator	ACME (95% CI)	p-value
Diabetes	815 (262, 1660)	<0.001
Dyslipidemia	-52 (-2030, 1830)	0.97
Hypertension	2170 (711, 4520)	<0.001
Smoking	-620 (-1960, 267)	0.18
BMI	1080 (92, 3330)	<0.001
PA	232 (-169, 867)	0.26

**Table 2.** Average causal mediation effects of modifiable risk factors on the association between education level and cardiovascular incidence at 6 years. ACMEs and their corresponding CIs are expressed as the expected difference in the number of CVD-free days when the mediator takes the value observed in individuals with university education compared to those with primary or lower education. Mediators were modeled with linear (BMI, PA), logistic (Diabetes, Dyslipidemia, Hypertension), and ordered logistic (Smoking) regression models; CVD incidence with a parametric survival model. Monte Carlo simulations were used to obtain the CIACME: average causal mediation effect; BMI: body mass index; CI: confidence interval; PA: physical activity.

modifiable risk factors such as diabetes, hypertension, and BMI. Further analyses showed that awareness, treatment, and control of diabetes and hypertension were not mediating the association between education and CVD incidence, and that hypertension was the strongest mediator.

Our results showed a reduction of CVD incidence in participants with university education compared to participants with primary or lower education. While the association between education and CVD incidence had not been assessed with causal methods before, our estimates are in line with others from population-based cohorts<sup>6,9,23,24</sup>. This result suggests that different prevention strategies or at least different intensities in prevention should be provided to individuals by educational level to account for the excess CVD risk and for the clustering of CVD risk factors in low socioeconomic groups<sup>25</sup>.

A graded inverse association between SES and CVD has been described in some studies<sup>13,14</sup>. We think that the lack of effect observed when comparing participants with secondary education with those with primary or lower education in the present analysis could be caused by a more heterogeneous secondary education group. This expectation is in accordance with the effect modification identified for age in the secondary education group. In addition, a possible birth cohort effect characterized by different percentages of the 3 educational categories in the 3 included cohorts could be playing a role.

Our results showed that diabetes, hypertension, and BMI, were mediators of the association between education and CVD incidence. These mediators accounted for 26% of the total effect of education on CVD incidence. In previous studies the effect of education on CHD incidence was attenuated up to a 70% after adjusting for CV and behavioral risk factors<sup>6,19,24</sup>. More recent studies focusing on the mediation of the association between education and CHD showed that 21–57% of the effect was accounted for behavioral and biological risk factors<sup>10,18,20</sup>. These and other studies pointed to smoking, BMI, hypertension, PA, lipid profile, and diabetes as mediators of the association between education and CVD incidence<sup>10,11,13,15,16,20,24,26</sup>.

From the 6 modifiable risk factors analyzed (diabetes, dyslipidemia, hypertension, smoking, BMI, and PA) only diabetes, hypertension, and BMI, were identified as mediators of the association between education and CVD incidence. There are no studies analyzing the mediators of this association using the counterfactual framework, however, a recent study by Kershaw *et al.*, which estimates indirect effects of classical and lifestyle risk factors simultaneously may be a fair comparison<sup>20</sup>. In this study, in addition to diabetes, hypertension, and BMI, they found that hypercholesterolemia, smoking, and PA were also mediators of the education-CHD relationship. Disagreement on the mediating effect of hypercholesterolemia and PA is probably influenced by differences in variable measurement. In the REGICOR cohorts PA was based on the Minnesota leisure-time physical activity questionnaire, cholesterol was measured in fasting samples and dyslipidemia was based on LDL-C levels. While Kershaw *et al.* obtained the PA information from a single question, measured cholesterol in non-fasting samples and based hypercholesterolemia in total cholesterol levels.

Our results did not show a significant mediation effect of smoking in the association between education and CVD incidence in contrast to the literature and to previous studies of mediation in this topic<sup>18–20</sup>. This lack of effect is probably influenced by several factors being probably the more important the differential effect of smoking on the association between education and CVD incidence among European populations. In a study of European cohorts it was found that the percentage change in the relative index of inequality regarding education for CHD incidence due to smoking differed between European countries<sup>10</sup>. This difference was observed in both men and women, and in women it showed a contrary sign in Southern and in Northern European populations.

The identification of body mass index, diabetes, and particularly hypertension, as mediators of the education-CVD association, point to these risk factors as potential targets for interventions to reduce CVD in low socioeconomic groups. While there is a scarcity of interventions to improve risk factor control in those with low literacy<sup>2</sup> the few existing studies show positive effects. An American program developed individual- and community-level approaches targeting CVD health behaviors for blacks in their communities<sup>27</sup>. The program was associated with significant improvements in diet among others determinants<sup>28</sup>. In addition, in a study from Iran it was found that an educational and environmental program was able to reduce adiposity measures in low-literate women<sup>29</sup>. As these examples suggest, current recommendations are to include a combination of both population- and individual-level strategies to lower CVD risk and to target high-risk individuals<sup>30</sup>.

While diabetes and hypertension had already been described as mediators of the education-CVD incidence association, no study had previously analyzed the robustness of this effect regarding the causal mediation assumptions, or the effect of awareness, treatment, and control of these risk factors. Our results showed that hypertension was the most robust mediator based on the effect of the education-risk factor interaction terms, which yielded almost equal estimates for the hypertension ACME. In addition, neither awareness, treatment, or control, mediated the education-CVD incidence association in participants with diabetes or hypertension. Although the ACMEs for diabetes awareness and treatment, and for hypertension awareness were close to significant. It is possible that the reduced number of individuals with diabetes or hypertension, 13% and 42%, respectively, was an insufficient sample size to examine the mediation effect of diabetes and hypertension awareness, treatment, and control.

Modifiable CV risk factors explained only 26% of the association between education and CVD incidence. We expect that other factors would explain part of the remaining effect. Diet could also act as a mediator on this association, but based on previous studies we believe this effect would be small once other modifiable risk factors have been accounted for<sup>20</sup>. Other mediators could be biological (e.g. stress biomarkers), material (e.g. financial issues), early life and psychological characteristics (e.g. material and cognitive childhood factors), or socio-economic, cultural or environmental factors as pointed out by some authors<sup>12,14,31,32</sup>. Psychological mechanisms are probably accounting for a part of the non-explained association between education and CVD as a large number of studies have described a strong relation between emotional states and CVD risk, particularly depression<sup>33</sup>. However, the scarcity of studies examining the mediating role of psychological factors precludes the estimation of this effect.

To our knowledge, this is the first study estimating ACMEs of multiple modifiable risk factors in the education-CVD incidence association. Our study is also the first addressing the role of modifiable risk factor awareness, treatment, and control. The main strengths are the use of a large population-based cohort ( $n = 9,226$ ), randomly selected from a reference population of 600,000 inhabitants, and novel statistical tools such as the counterfactual framework.

There are some limitations that should be considered though. First, due to limitations on the available analytical tools we could not assess the confounding effect of some mediators (e.g. BMI) on the effect of other mediators (e.g. diabetes and hypertension). Another limitation is the assumption of no unmeasured confounding. While this assumption cannot be tested, its plausibility can be assessed with sensitivity analyses. Our sensitivity analyses showed that unmeasured confounding seem not to be affecting the estimates. Finally, we did not include medication adherence and other socioeconomic factors which could explain part of the association between education and CVD incidence and affect diabetes and hypertension control as they were not available in the REGICOR cohorts.

In conclusion, individuals from North-Eastern Spain with a university degree had a CV event HR of 0.51 (95% CI = 0.30–0.85) at 6 years, compared to the ones who received primary or lower education. The association between education and CVD incidence was partially mediated by diabetes, hypertension, and BMI, being hypertension the strongest mediator. These mediating factors, particularly, hypertension, could be tackled to diminish CVD risk among population with a low education level.

## Methods

**Study population.** The Registre Gironí del Cor (REGICOR) cohorts, were recruited in three population-based surveys conducted in the province of Girona (north-eastern Spain) in 1995 ( $n = 1748$ ), 2000 ( $n = 3058$ ) and 2005 ( $n = 6352$ )<sup>21</sup>. At baseline, participants underwent a physical examination and completed interview-administered questionnaires<sup>22</sup>. Participants were followed for cardiovascular (CV) events until 2011–2013 (median follow-up 7 years). This study included participants from the REGICOR cohorts aged 35–74 years with no history of CVD (acute myocardial infarction -AMI- or stroke), and with available data in the variables of interest ( $n = 9226$ ) (Fig. 1).

REGICOR surveys were approved by the Parc de Salut Mar ethics committee and informed consent was obtained from all participants. All experiments were performed according to guidelines and regulations.

**Exposure and outcome.** The exposure under study was education level attained, reported by the participants at baseline. Education was categorized in three levels: primary or lower education (any education below the secondary school diploma), secondary (secondary school diploma or higher without a college degree) and university (college degree).

The outcome of interest was CVD incidence during the first 6 years of follow-up analyzed as a time-to-event variable. CVD events were identified by participant interviews, medical record revision, and data linkage with the REGICOR AMI population registry<sup>34</sup> and the official mortality register. The REGICOR AMI registry registered all AMI that occurred in 1977–2009 in the same region where the population cohorts were recruited. Fatal events were identified using ICD9 codes 410–414 and ICD10 codes I20–I25 for coronary events, and ICD9 codes 430–438 and ICD10 codes I60–I69 for cerebrovascular events. Non-fatal cases were identified from medical records and physician notes, and were classified in committee according to standardized criteria: i) AMI was defined according to symptoms, electrocardiogram, and biomarkers of necrosis; ii) angina was defined according to the presence of symptoms and objective demonstration of ischemia on ECG or presence of coronary stenosis; iii) coronary revascularization, including percutaneous invasive revascularization and surgery; iv) stroke was defined based on objective evidence of cerebral ischemic injury (pathological, imaging, or other) in a defined vascular distribution, or clinical evidence of cerebral focal ischemic injury based on symptoms persisting  $\geq 24$  hours or until death, and other etiologies excluded<sup>35</sup>.

**Covariates.** Descriptive analyses by education level included sociodemographic variables (age and gender), CV risk factors (diabetes, hypertension, dyslipidemia, smoking, BMI, and PA), social class, and quality of life. CV risk factors were recorded at baseline as previously described<sup>21,22</sup>. Presence of diabetes, hypertension, and dyslipidemia were based on previous diagnosis, current treatment and baseline values of fasting glucose, blood pressure and low density lipoprotein cholesterol (LDL-C). The cutoff for fasting glucose was  $\geq 126$  mg/dL, for systolic and diastolic blood pressure was  $\geq 140$  and  $\geq 90$  mmHg, respectively, and for LDL-C was  $\geq 160$  mg/dL. Among participants with diabetes or hypertension we also analyzed the awareness (proportion who declared to be aware of having the risk factor), treatment (proportion treated), and control (proportion with glucose or blood pressure under the defined cutoffs). Smoking was categorized in non-smokers, ex-smokers (more than 1 year), and smokers (at least 1 cigarette/day). We calculated BMI as weight divided by squared height ( $\text{kg}/\text{m}^2$ ). Individuals were classified as inactive if they spent  $< 600$  metabolic equivalents (METs)/week of moderate PA and  $< 400$  METs/week of vigorous PA, and as active if they did  $\geq 600$  METs/week of moderate PA or  $\geq 400$  METs/week of vigorous PA<sup>36</sup>, based on the Minnesota leisure-time physical activity questionnaire. Social class was categorized in manual and non-manual based on occupation as recommended by the Spanish Society of Epidemiology<sup>37</sup>. Quality of life physical and mental summary components were obtained from the Spanish version of the SF-36 Health survey<sup>38</sup>.

**Statistical analysis.** A directed acyclic graph for the causal association of education and CVD was created using DAGitty<sup>39</sup> based on previous research experience and literature evidence (Suppl. Fig. 2).

Continuous variables were summarized as mean and standard deviation, and categorical variables as proportions. P-value for trend was computed using the Pearson test.

The association between education and CVD incidence was examined with a marginal structural survival model for the outcome time to CVD event in the first 6 years of follow-up, using inverse-probability-of-treatment-and-censoring (IPTC) weights and robust standard errors. The proportional hazard assumption was confirmed graphically and by the nonzero slope test. IPTC weights were calculated as the product of stabilized exposure of interest (education level) and censoring weights<sup>40</sup>. For exposure weights, the denominator was each participant's probability of having their education level based on all measured pre-education covariates (age, gender and cohort), while the numerator was this probability based on the intercept and no covariates. Age was included as a smoothing spline using a penalized spline basis. Cohort was included as a 3 category variable. Multinomial logistic regression was used to create the weight models for exposure. For censoring weights, the denominator was each participant's probability of being censored based on age, gender, cohort and education, while the numerator was this probability based on education only. Logistic regression was used to create the weight models for censoring. Effect modification by age and gender was analyzed by adding these covariates plus their interaction with education to the structural model.

Assumptions of the marginal structural model were analyzed as follows<sup>41</sup>. Exchangeability was examined by analyzing the effect of unmeasured categorical confounders. First, the inclusion of place of birth in the weights was analyzed. Place of birth was included as a 3 category variable (Catalonia, north region of the rest of Spain, and south region of the rest of Spain). Second, the obsSens package from R was used to analyze potential unknown confounders. To analyze positivity, we re-fitted the models excluding the subgroup where there was a low number of participants with university education (age  $> 70$  years). Model specification was assessed by using different models (survival and logistic), by excluding participants from cohort 2000, and by redoing the analysis with different follow-up times (4, 5, 7 and 8 years).

Mediation of the association between education level and 6-year CVD incidence by modifiable risk factors (diabetes, hypertension, dyslipidemia, smoking, BMI, and PA) was analyzed with the counterfactual framework using the R package mediation<sup>42</sup>. The mediation effect of awareness, treatment, and control of the significant treatable mediators was also examined in individuals with each of the risk factors. The counterfactual framework allows the estimation of indirect effects of a treatment on an outcome through a set of mediators based on models for the mediators and the outcome. Mediator models were linear (BMI, and PA), logistic (diabetes, dyslipidemia, and hypertension), or ordered logistic (smoking) depending on the mediator. The outcome (6-year CVD incidence) was modeled using a parametric survival model. Quasi-bayesian confidence intervals were calculated for the estimates using 1000 simulations. Age, gender and cohort were included as pre-treatment confounders. We calculated the percentage of mediated association as the risk factor average causal mediation effect (indirect effect) divided by the total effect of education on CVD incidence<sup>20</sup>.

Assumptions of the mediation analysis were examined as follows<sup>43</sup>. The assumption of temporal ordering was satisfied as the exposure (education level attained prior to baseline) preceded the outcome (CVD incidence during follow-up), mediators –measured at baseline– preceded the outcome –measured from baseline on–, and exposure preceded the mediators. Unmeasured confounding was examined with a sensitivity analysis which included place of birth and social class as confounders in the mediator and outcome models. Measurement error of the mediators was assessed for PA by treating this variable as continuous and categorical in separate models. Estimate bias due to interaction between exposure and mediators was assessed in sensitivity analysis by including this interaction in the outcome models and checking the change in the indirect effect estimates.

Analyses were conducted in R statistical software version 3.2.3<sup>44</sup>.

**Data Availability.** The dataset analyzed during the current study is available from the corresponding authors on reasonable request.

## References

- World Health Organization. The determinants of health. <http://www.who.int/hia/evidence/doh/en/>. Accessed on 07/07/ (2017).
- Havranek, E. P. *et al.* Social Determinants of Risk and Outcomes for Cardiovascular Disease: A Scientific Statement From the American Heart Association. *Circulation*. **132**, 873–898 (2015).
- Meara, E. R., Richards, S. & Cutler, D. M. The gap gets bigger: changes in mortality and life expectancy, by education, 1981–2000. *Health Aff (Millwood)*. **27**, 350–360 (2008).
- Rose, G. & Marmot, M. G. Social class and coronary heart disease. *Br Heart J*. **45**, 13–19 (1981).
- Emberson, J. R. *et al.* Social class differences in coronary heart disease in middle-aged British men: implications for prevention. *Int J Epidemiol*. **33**, 289–296 (2004).
- Silventoinen, K. *et al.* Educational inequalities in the metabolic syndrome and coronary heart disease among middle-aged men and women. *Int J Epidemiol*. **34**, 327–334 (2005).
- Avendano, M. *et al.* Socioeconomic status and ischaemic heart disease mortality in 10 western European populations during the 1990s. *Heart*. **92**, 461–467 (2006).
- Kuper, H. *et al.* The socioeconomic gradient in the incidence of stroke: a prospective study in middle-aged women in Sweden. *Stroke*. **38**, 27–33 (2007).
- Ernsten, L., Bjerkeset, O. & Krokstad, S. Educational inequalities in ischaemic heart disease mortality in 44,000 Norwegian women and men: the influence of psychosocial and behavioural factors. The HUNT study. *Scand J Public Health*. **38**, 678–685 (2010).
- Veronesi, G. *et al.* Educational class inequalities in the incidence of coronary heart disease in Europe. *Heart*. **102**, 958–965 (2016).
- Lynch, J. W. *et al.* Do cardiovascular risk factors explain the relation between socioeconomic status, risk of all-cause mortality, cardiovascular mortality, and acute myocardial infarction? *Am J Epidemiol*. **144**, 934–942 (1996).
- Kamphuis, C. B. M. *et al.* Socioeconomic inequalities in cardiovascular mortality and the role of childhood risk factors: a prospective cohort study with 17-years of follow up. *BMC Public Health*. **12**, 1045 (2012).
- Kuper, H. *et al.* Psychosocial determinants of coronary heart disease in middle-aged women: a prospective study in Sweden. *Am J Epidemiol*. **164**, 349–357 (2006).
- Ramsay, S. E. *et al.* Socioeconomic inequalities in coronary heart disease risk in older age: contribution of established and novel coronary risk factors. *J Thromb Haemos*. **7**, 1779–1786 (2009).
- Marmot, M. G. *et al.* Biological and behavioural explanations of social inequalities in coronary heart disease: Whitehall II Study. *Diabetologia*. **51**, 1980–1988 (2008).
- Strand, B. H. & Tverdal, A. Can cardiovascular risk factors and lifestyle explain the educational inequalities in mortality from ischaemic heart disease and from other heart diseases? 26 year follow up of 50,000 Norwegian men and women. *J Epidemiol Community Health*. **58**, 705–709 (2004).
- Kivimäki, M. *et al.* Socioeconomic position, co-occurrence of behavior-related risk factors, and coronary heart disease: the finnish public sector study. *Am J Public Health*. **97**, 874–879 (2007).
- Nordahl, H. *et al.* Education and risk of coronary heart disease: assessment of mediation by behavioral risk factors using the additive hazards model. *Eur J Epidemiol*. **28**, 149–157 (2013).
- Méjean, C. *et al.* The contribution of diet and lifestyle to socioeconomic inequalities in cardiovascular morbidity and mortality. *Int J Cardiol*. **168**, 5190–5195 (2013).
- Kershaw, K. N. *et al.* Quantifying the contributions of behavioral and biological risk factors to socioeconomic disparities in coronary heart disease incidence: The MORGEN study. *Eur J Epidemiol*. **28**, 807–814 (2013).
- Grau, M. *et al.* Trends in cardiovascular risk factors prevalence (1995–2000–2005) in northeastern Spain. *Eur J Cardiovasc Prev Rehabil*. **14**, 653–659 (2007).
- Redondo, A. *et al.* Trends in the prevalence, awareness, treatment, and control of cardiovascular risk factors across educational level in the 1995–2005 period. *Ann Epidemiol*. **21**, 555–563 (2011).
- Veronesi, G. *et al.* Gender differences in the association between education and the incidence of cardiovascular events in Northern Italy. *Eur J Public Health*. **21**, 762–767 (2011).
- Van Lenthe, F. J. *et al.* Material and behavioural factors in the explanation of educational differences in incidence of acute myocardial infarction: the globe study. *Ann Epidemiol*. **12**, 535–542 (2002).
- Marmot, M. & Wilkinson, R. *Social Determinants of Health*. Oxford, UK:Oxford University Press; 2006:80.
- Floud, S. *et al.* The role of health-related behavioural factors in accounting for inequalities in coronary heart disease risk by education and area deprivation: prospective study of 1.2 million UK women. *BMC Medicine*. **14**, 145 (2016).
- Plescia, M., Groblewski, M. & Chavis, L. A lay health advisor program to promote community capacity and change among change agents. *Health Promot Pract*. **9**, 434–439 (2008).
- Plescia, M., Herrick, H. & Chavis, L. Improving health behaviors in an African American community: the Charlotte Racial and Ethnic Approaches to Community Health project. *Am J Public Health*. **98**, 1678–1684 (2008).
- Sadeghi, M. *et al.* Women health heart project: Methodology and effect of interventional strategies on low education participants. *J Educ Health Promot*. **28**, 3–103 (2014).
- Rose, G. Sick individuals and sick populations. *Int J Epidemiol*. **14**, 32–38 (1985).
- Loucks, E. B. *et al.* Education and coronary heart disease risk associations may be affected by early-life common prior causes: a propensity score matching analysis. *Ann Epidemiol*. **22**, 221–232 (2012).
- Falkstedt, D. & Hemmingsson, T. Educational level and coronary heart disease: a study of potential confounding from factors in childhood and adolescence based on the Swedich 1969 conscription cohort. *Ann Epidemiol*. **21**, 336–342 (2011).
- Barth, J., Schumacher, M. & Herrmann-Lingen, C. Depression as a risk factor for mortality in patients with coronary heart disease: a metaanalysis. *Psychosom Med*. **66**, 802–813 (2004).
- Marrugat, J. *et al.* REGICOR Investigators. Acute myocardial infarction population incidence and in-hospital management factors associated to 28-day case-fatality in the 65 year and older. *Eur J Epidemiol*. **19**, 231–237 (2004).
- Sacco, R. L. *et al.* An updated definition of stroke for the 21<sup>st</sup> century: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*. **44**, 2064–2089 (2013).
- Haskell, W. L. *et al.* American College of Sports Medicine; American Heart Association. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Circulation*. **116**, 1081–1093 (2007).
- Grupo de Trabajo de la SEE – Alvarez-Dardet, C., Alonso, J., Domingo, A. & Regidor, E. La medición de la clase social en ciencias de la salud (SG Editores, 1995).
- Alonso, J., Preto, L. & Anto, J. M. The Spanish version of the SF-36 Health Survey (the SF-36 health questionnaire): an instrument for measuring clinical results. *Med Clin (Barc)*. **104**, 771–776 (1995).
- Textor, J., Hardt, J. & Knüppel, S. DAGitty: A Graphical Tool for Analyzing Causal Diagrams. *Epidemiology*. **5**, 745 (2011).
- Hernán, M. A. & Robins, J. M. Causal Inference. Boca Raton: Chapman & Hall/CRC, forthcoming (2017).
- Cole, S. R. & Hernán, M. A. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol*. **168**, 656–664 (2008).

42. Tingley, D. *et al.* Mediation: R Package for Causal Mediation Analysis. *J Stat Softw.* **59**, 1–38 (2014).
43. VanderWeele, T. J. Unmeasured confounding and hazard scales: sensitivity analysis for total, direct, and indirect effects. *Eur J Epidemiol.* **28**, 113–117 (2013).
44. R Core TeamR: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/> (2017).

## Acknowledgements

This work was supported by the RECERCAIXA Program, Obra Social “La Caixa” (RE087465). The REGICOR cohorts were funded by the Spanish Ministry of Economy through the Carlos III Health Institute (Red de Investigación Cardiovascular Programa HERACLES RD12/0042, Red RedIAPP RD06/0018, CIBERCV CB16/11/00229), European Regional Development Funds (ERDF-FEDER) Health Research Fund (FIS-90/0672, FIS-93/0568, FIS 94/0539, FIS96/0026-01, FIS99/0655, FIS99/0013-01, FIS 99/9342, FIS 2003/HERMES, PI20471, FIS14/00449, CP12/03287); the Government of Catalunya through Agència de Gestió Ajuts Universitaris de Recerca [2014 SGR 240]. The authors would like to thank Susanna Tello, Marta Cabañero and Leny Franco for data management, and Anna Puigdefàbregas and Rosa Gispert from the Catalan Government for providing cross-linkage of our databases with the Mortality Registry. The authors would also like to thank all participants and staff (MDs, nurses, researchers and technicians) from the REGICOR cohorts.

## Author Contributions

I.R.D. and R.E. designed the study. R.R., A.Z. and R.M. participated in data obtaining. R.E. and J.M. provided the data and supervised the analysis. I.R.D. did the statistical analysis. I.R.D. and R.E. analyzed the results. J.M., M.G., and B.S. provided comments and suggestions during data analysis. I.R.D. drafted the manuscript. All authors revised the manuscript for important intellectual content and approved the final version.

## Additional Information

**Supplementary information** accompanies this paper at doi:[10.1038/s41598-017-10775-3](https://doi.org/10.1038/s41598-017-10775-3)

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017