Education and health & well-being: Direct and indirect effects with multiple mediators and interactions with multiple imputed data in Stata

Mashhood Ahmed Sheikh*, Birgit Abelsen, Jan Abel Olsen

Citation:

Sheikh, M. A., Abelsen, B., & Olsen, J. A. (2017). Education and health and well-being: direct and indirect effects with multiple mediators and interactions with multiple imputed data in Stata. *Journal of Epidemiology and Community Health, 71*(11), 1037-1045. doi:10.1136/jech-2016-208671

Health Services Research Unit, Department of Community Medicine, University of Tromsø, NO 9037 Tromsø, Norway.

Corresponding author: Mashhood Ahmed Sheikh, Department of Community Medicine, University of Tromsø, 9037 Tromsø, Norway; Tel.: 0047-77620716; E-mail: mashhood.a.sheikh@uit.no.

Email addresses:

MAS: mashhood.a.sheikh@uit.no

BA: birgit.abelsen@uit.no

JAO: jan.abel.olsen@uit.no

Authors' contributions: MAS developed the theory, analytic strategy, performed statistical analysis, data interpretation, and wrote the manuscript. BA and JAO conceived the substantive research questions and contributed in revising the manuscript. All authors have read and approved the final manuscript.

Running head: Education and health & well-being

Funding: This research was funded by UiT The Arctic University of Norway

Competing Interest: None declared.

What is already known on this subject?

- Previous methods for assessing mediation rely on ‘no interaction’ assumptions between the exposure, mediator(s), and covariates.
- Education, income, management position, occupational hierarchy position, and subjective social status are associated with health and well-being.
- Indicators of adult socioeconomic status partially or wholly mediate the association between education and health.

What this study adds?

- The Inverse Odds Weighting approach, which does not rely on ‘no interaction’ assumptions, was used to assess mediation.
- We provide Stata code that makes it easy to assess mediation by the Inverse Odds Weighting approach in any multiple imputed dataset.
- Education influences health and well-being through different pathways.

ABSTRACT

Background: Previous methods for assessing mediation assume no multiplicative interactions. The Inverse Odds Weighting (IOW) approach has been presented as a method that can be used even when interactions exist. The substantive aim of this study was to assess the indirect effect of education on health and well-being via four indicators of adult socioeconomic status (SES): income, management position, occupational hierarchy position, and subjective social status. **Methods:** 8,516 men and women from the Tromsø Study (Norway) were followed for 17 years. Education was measured at age 25-74 years, while SES and health and well-being were measured at age 42-91 years. Natural direct (NDE) and indirect effects (NIE) were estimated using weighted Poisson regression models with IOW. Stata code is provided that makes it easy to assess mediation in any multiple imputed dataset with multiple mediators and interactions. **Results:** Low education was associated with lower SES. Consequently, low SES was associated with being unhealthy and having a low level of well-being. The effect (NIE) of education on health and well-being is mediated by income, management position, occupational hierarchy position, and subjective social status. **Conclusion:** This study contributes to the literature on mediation analysis, as well as the literature on the importance of education for health-related quality of life and subjective well-being. The influence of education on health and well-being had different pathways in this Norwegian sample.

Keywords: Education, Subjective Social Status, Income, Occupational hierarchy position, Social Epidemiology, Health-related quality of life, Subjective well-being

INTRODUCTION

Previous approaches (e.g., the product-of-coefficients method and the difference-in-coefficients method, including those based on counterfactuals/potential outcomes) for identifying and estimating unbiased direct and indirect effects rely on the following assumptions [1-13].

1. Temporality between the exposure, mediator(s), and outcome.
2. No measurement error in the exposure, mediator(s), or outcome.
3. No unmeasured (or unaccounted-for) confounding of the exposure-mediator, exposure-outcome, or mediator-outcome associations.
4. No exposure-induced mediator-outcome confounding, i.e., no variables (measured or unmeasured) that affect both the mediator(s) and the outcome that are affected by the exposure itself.
5. Correct form and specification of the regression models for exposure-mediator, exposure-outcome, and mediator-outcome association (e.g., linear, fractional polynomial, etc.)
6. No exposure-mediator, exposure-covariate, mediator-mediator, or mediator-covariate multiplicative interaction.

Note that assumption 1 (temporality between exposure and outcome), assumption 2 (no measurement error in exposure and outcome), assumption 3 (no unmeasured confounding of the exposure-outcome association), assumption 5 (correct form of the regression model for exposure-outcome association), and assumption 6 (no exposure-covariate interaction) are neither novel nor new to applied researchers, as these assumptions are also employed when assessing ‘unbiased’ associations between an exposure and an outcome. The only addition is the inclusion of a ‘mediator’ in these assumptions. For practical purposes, we divided these six assumptions into two categories, fixable assumptions (FA) and unfixable assumptions (UA). The terms fixable and unfixable refer to statistical solutions at the analysis stage of the research. FA are those that a researcher can tackle statistically after collecting the data, while UA are those that the researcher can try to address, but can never be 100% certain they are not a concern when interpreting estimates. It is important to distinguish between UA and FA, because UA require intervention at the design stage of the research (i.e., prior to data collection), while FA can be tackled at the analysis stage (i.e., after data collection).

Assumptions 1, 2, 3, and 4 are UA, while assumptions 5 and 6 are FA. Assumption 1 is UA since temporality cannot be introduced in the data with a statistical method. For some research questions, such as those related to a strong theoretical framework like the social causation hypothesis, temporality is often assumed, but cannot be ascertained in cross-sectional data [14, 15]. Assumption 2 is also UA after collecting the data. For example, attempts can be made for controlling 'unreliability' either by using the reliability coefficient (if available) [16], or by constructing a latent variable within the structural equation modelling framework [17, 18]. However, the unreliability cannot be controlled with 100% certainty despite these statistical 'fixes'. The possibility of differential measurement error complicates assumption 2, as the differential and non-differential measurement error may cancel each other out to some extent [19]. Indeed, even if all the confounding variables are measured and included in the models, it is still important that all confounding variables be measured without error, and that they be specified correctly in the regression models.

If both the exposure and mediator are randomly assigned at the design stage, than assumption 3 and 4 are automatically taken care of; however, the exposure can not affect the mediator. Consequently, mediation (indirect effect) cannot be assessed. If the exposure is the only randomly-assigned variable, then assumption 3 regarding exposure-mediator and exposure-outcome confounding are taken care of; however, unmeasured or unaccounted-for mediator-outcome confounding would remain a concern. Post-design strategies, such as propensity score matching, may help tackle unmeasured confounding assumptions, but do not eliminate the threat to the validity of estimates. Similarly, several strategies [20-26] for tackling assumption 4, no intermediate confounders, have been proposed, but the possibility of unmeasured intermediate confounding remains a threat to the identification and interpretation of direct effect estimates, because there may be numerous unmeasured intermediate confounding variables with unknown distributions and unpredictable strength and associations between themselves and with the covariates, exposure, mediator(s), and outcome.

Others have proposed minimizing the time period between the exposure and mediator. For example, recent studies [11, 12] assessed the mediating role of childhood abuse in the association between childhood socioeconomic status (SES) and health and well-being in adulthood. Such study designs may reduce, but cannot rule out the possibility of intermediate confounding. While some UA can be partially addressed in the analysis stage (as discussed above), the best solutions are implemented at the design stage of the project. Assumption 5 can be tackled by carefully analyzing the form of the association between the exposure, mediator(s), and outcome, and choosing appropriate statistical models that fit the data [13].

A potential solution for assumption 6 can be explained by referring to an empirical example. The aim of the empirical example used in this study was to assess the direct and indirect effect of education on health and well-being in adulthood, using a population-based cohort study in Norway. Several studies have shown that education is associated with better health and well-being, but it is unclear exactly how [27, 28]. Elder [29] proposed the life course perspective as pathways or trajectories defined by sequences of life conditions and social standing. Thus, an education level trajectory can be defined by upward or downward social mobility (via income, management position, occupational hierarchy position, and subjective social status), and its influence on health and well-being in adulthood. Lower education level may lead to exposure to stress from a range of sources [30, 31], which may ultimately affect health and well-being. For instance, respondents with a low level of education are more likely to have an occupation with lower income and prestige [32]. Previous studies have shown that lower income and low job control contribute to lower health and well-being [33-35]. On the contrary, a higher education has been suggested to protect against exposure to negative life events and chronic stressors, in part through an increased sense of control over life events through effort and action [36-38].

Education could affect health and well-being through increased income; power and control at work (management position); prestige and nobility (occupational hierarchy position); perceived higher ranking in a social hierarchy (subjective social status); or any combination of these. Adult SES can be measured by any of the aforementioned indicators [39], and previous research has shown that income [39-41], management position [42], occupation type [39, 41, 42], and subjective social status [43, 44] are all associated with health and well-being.

The assessment of the indirect pathways between education and health and well-being is limited by the statistical methods available in the literature. Assumption 6 concerns interactions between the exposure and mediator(s), the exposure and covariate (s), between different mediators, and between the mediator and covariate(s), regressed on the outcome. Previous studies [39-49] have suffered from methodological limitations, as they did not account for the interactions between education and SES, between indicators of SES, and between SES and confounding variables. Indeed, assessing mediation without accounting for interactions leads to biased estimates [11, 50]. An 'easy way out' is to simply not test the interactions, and there has been no solution for this until recently.

Disentangling the pathways between education and health and well-being requires a statistical methodology that allows interactions between indicators of SES, between exposure and mediator(s), and between SES and confounding variables. Previous statistical approaches to assess mediation assumed no interaction between the exposure and mediator(s) [1, 8-10, 51-53], and no interaction between mediators [5]. While those approaches may work when there are multiple causally independent mechanisms (no association, and no interaction between mediators), they are not applicable in settings where interactions are present. Since indicators of SES may interact, and co-occur [39] in the same individual, the 'no interaction' assumption may not be satisfied. Other approaches [54] are applicable only if the outcome is rare. However, if a high proportion of people classify themselves as 'unhealthy' or as having a 'low level of well-being', the assumption of a 'rare' outcome would not be satisfied.

Furthermore, several statistical approaches [55-58] to assess mediation do not allow interactions between mediator(s) and confounding variables. The effect of SES on health and well-being may depend on socio-demographic variables [14], and stratifying the analysis for each subgroup may not be feasible or meaningful. One common stratification is gender [11, 14], but the influence of SES on health and well-being may also depend on age group (cohort effect [59]) and childhood SES [14, 60, 61]. Other statistical approaches [62] that allow multiple mediators and interactions between mediators estimate direct and indirect effects separately for a full factorial combinations of mediators. This may not be practical or informative when there are more than two mediators with several categories, as the number of direct/indirect effects increases exponentially.

Recently, Tchetgen Tchetgen [63] proposed a method for estimating natural direct and indirect effects using inverse odds weights (IOW). This method is entirely agnostic with regard to interactions between mediators, between mediator(s) and exposures, or between mediator(s) and confounding variables [63, 64]. The method is valid regardless of whether such interactions are present, without having to assess and specify them [63, 64]. Thus, it provides a practical advantage over previous methods, in that the estimation of only one direct/indirect effect is sufficient for each mediator.

Another issue is the missing values in the dataset. One solution is to exclude all respondents with missing values on any variables in the analysis, but this can reduce sample sizes considerably. For example, an earlier study [14] assessed the direct and indirect effect (via education level) of childhood SES on health and well-being in adulthood, but had to discard almost 40% of the data due to missing values. For this reason, the results represented a highly selected group of participants that responded to all the questions included in the analysis. Any inference drawn from such a selected sample may not apply to the population the sample came from, let alone be comparable with any other population. Multiple imputation (MI) with chained equations and full information maximum likelihood have been considered the best strategies to address missing values. However, previous Stata software [21, 65-74] and other Stata routines (-ml_mediation-, -resboot_mediation-, -sgmediation-, and -binary_mediation-) for assessing mediation do not support MI data. Therefore we are providing here the Stata code that can be used with any MI dataset in Stata.

Using the IOW approach, the substantive aim of this study was to assess the mediating role of four indicators of SES (income, management position, occupational hierarchy position, and subjective social status) in the association between education and health and well-being.

METHODS

Study population

Tromsø is the largest city in Northern Norway, with more than 70,000 inhabitants. The Tromsø Study is a prospective cohort study of adult respondents residing in the municipality of Tromsø. It is considered representative of its adult population [75]. Six waves of the Tromsø Study, referred to as Tromsø I-VI, were conducted between 1974 and 2007/2010 [75]. The current paper is based on 8,516 respondents who participated in the Tromsø IV (1994-1995), and Tromsø VI (2007-2010) studies (aged 25-74 at baseline, and aged 42-91 at last measurement).

Data sources

Health and well-being. Measures of health and well-being were taken from Tromsø VI. Health was measured by the EQ-5D generic descriptive system for health-related quality of life. The EQ-5D includes five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension has three levels: 1) no problems, 2) some problems, and 3) unable or extreme problems. In the present analysis, all respondents who ticked level 1 for all five health dimensions were classified as healthy, and all others as relatively unhealthy [14, 15]. Subjective well-being was measured by the response to the first three items on the satisfaction with life scale (SWLS). These were: ‘In most ways my life is close to my ideal’; ‘The conditions of my life are excellent’, and; ‘I am satisfied with my life’. Respondents rated these statements using a 7-point scale ranging from completely disagree (1) to completely agree (7). We classified all those who ticked 6 or 7 on all three items as having a high level of subjective well-being, and all others as having relatively low level of subjective well-being. The cut-off points for classifying health and subjective well-being are in agreement with previous studies [14, 15].

Indicators of SES. Information on education was taken from Tromsø IV study (1994-1995), while other SES indicators were taken from Tromsø VI study (2007-2010). Respondents indicated their highest completed level of education: primary and secondary school or similar (i.e., 7–10 years of schooling); vocational school, high school, college, or university (less than 4 years); and college or university (4 years or more). All those who reported no college or university were classified as having low education and the remaining as having high education.

The question on income read 'What was the household's total taxable income last year?': 1) less than Norwegian Kroner (NOK) 125,000; 2) NOK 125,000-200,000; 3) NOK 201,000-300,000; 4) NOK 301,000-400,000; 5) NOK 401,000-550,000; 6) NOK 551,000-700,000; 7) NOK 701,000-850,000; and 8) more than NOK 850,000 (€1=NOK 9). Those who ticked options 1-4 were considered as having low income, those ticking 5-6 as having medium income, and those ticking 7-8 as having high income.

Management position was measured using two questions: 'Do you have management responsibilities in your position?', and; 'If yes, which kind of management responsibilities do you have in your position?: 1) work management, 2) middle management, 3) senior management, 4) project management'. Those responded 'no' to the first question were considered as having no management position. Those who replied yes to the first question and ticked the first two categories of the second question were considered as having a low-level management position, and those who ticked the last two categories of the second question were considered as having a high-level management position.

Occupational hierarchy position was measured as: 'Which of the following fields describe your profession?: 1) professions without education requirement; 2) process and machine operators or transport work; 3) artisan or similar;) agriculture, forestry, or fishery work; 5) sales, service, and care professions; 6) office and customer relations; 7) professions that require up to 3 years of university education; 8) professions that require at least 4 years of university education; 9) administrative leader or politician'. Those ticking any of categories 1-4 were classified as having a low occupational hierarchy position; those ticking categories 5-7 having a medium position, and; those ticking 8-9 were classified as having a high occupational hierarchy position.

Subjective social status was measured as: 'I consider my occupation to have the following social status in the society (if you are not currently employed, think about your recent occupation)'. The response alternatives were: 1) very high; 2) high; 3) neither high nor low; 4) low, and; 5) very low. Those ticking 1-2 were classified as having a high subjective social status, while those ticking 4-5 were classified as having a low subjective social status. Subjective social status is expected to be adequately reliable [76].

Potential confounding variables. Data on the potential confounding variables were taken from the Tromsø IV questionnaire: age, gender, childhood SES (very good or good, difficult, or very difficult), exposure to passive smoking in childhood (yes, no), having enough friends (yes, no), number of friends (0-3, 4-5, 6 or more), marital status (single, married or registered partnership, widow/widower, divorced or separated), and physical activity in hours/week (less than 1, 1-2, 3 or more). These confounding variables were chosen based on a *priori* knowledge of the association between the exposure and mediators, mediators and outcomes, and exposure and outcomes under study [13-15].

Statistical analysis

All analyses were conducted using Stata version 13. To avoid a reduction in sample size and bias related to list-wise deletion, missing values were imputed by the MI procedure in Stata version 13. Missing values were imputed using 50 MI datasets. The proportion (%) of respondents in the unimputed dataset, and in the MI dataset is presented (Table 1).

We did not observe any statistically significant interactions with gender. Therefore, all estimates are presented for men and women combined. The associations between education and adult SES, education and health and well-being, and SES and health and well-being were assessed with cross-tabulation and chi-square tests (Table 2). Associations between indicators of SES were assessed with the Pearson correlation, and 95% confidence intervals (CIs) were calculated with Fisher's z' transformation (see eTable 1).

Our aim was to estimate the natural direct and indirect effects of education on health and well-being in adulthood. The mediators were selected based on prior theory and the Causal Steps method [9, 10, 13-15]. We used the difference-in-coefficients method with IOW [63, 64] to assess mediation. An odds ratio scale is not suitable for assessing mediation, as this tends to overestimate the direct effect and underestimate the indirect effect, particularly when the outcome is not rare (e.g., proportion of unhealthy >5%) [15]. This bias in direct/indirect effects increases with prevalence of the outcome; therefore, we recommend estimating relative risks (RRs), even when the outcome is rare. For this reason, we used Poisson regression analysis to estimate the RR and 95% CI of being unhealthy and having a low level of well-being in adulthood dependent on education.

The algorithm for estimating total effects (RR_{TE}), natural direct effects (RR_{NDE}), and natural indirect effects (RR_{NIE}) was carried out in the following steps:

1. Fit the model for MI using mi impute chained software in Stata version 13.
2. Fit a logistic regression model for education (0=high, 1=low) conditional on the indicators of SES and confounding variables.
3. Compute an IOW by taking the inverse of the predicted log odds from step 2 for each observation in the exposed group (education=1 (low)).
4. Assign the IOW of each observation in the unexposed group (education= 0 (high)) equal to 1.
5. Estimate the natural direct effect of education via weighted generalized linear model (family=Poisson) of the regression of the outcome on education and confounding variables, with link=log function and the weights obtained in steps 3 and 4.

6. Estimate the total effect of education using a generalized linear model (family=Poisson) of the regression of the outcome on education and confounding variables, with link=log function.
7. Calculate the natural indirect effects of education on the outcome(s) via the proposed mediator(s) by subtracting the natural direct effects from the total effects as:

$$\beta_{\text{Natural indirect effect}} = \beta_{\text{Total effect}} - \beta_{\text{Natural direct effect}}$$

8. Bootstrap (5000 replications) with the bias-corrected accelerated bootstrap method to derive standard errors for total effect, natural direct effect, and natural indirect effect.

Steps 2-8 were followed for all mediators combined (Table 3), and separately for each mediator (Table 4). The Stata code for estimating total effect, natural direct effect, and natural indirect effect is provided in eAppendix. In addition, mediation was assessed with the traditional difference method approach, and resultant estimates are provided for comparison purposes in Table 3 and Table 4.

RESULTS

The distribution of general characteristics in the study sample was similar in the unimputed dataset and in the MI dataset (Table 1). The tests for linear trend showed that SES was associated with education in the expected direction; people who reported low education had lower SES ($p<0.05$). Similarly, having low education was associated with being unhealthy and having a low level of well-being ($p<0.05$). Income, management position, occupational hierarchy position, and subjective social status were significantly ($p<0.05$) correlated with each other (see eTable 1). The strongest correlation was between income and occupational hierarchy position (r: 0.43, 95% CI: 0.41, 0.44), while the weakest correlation was between income and management position (r: 0.22, 95% CI: 0.20, 0.24), and between income and subjective social status (r: 0.22, 95% CI: 0.20, 0.24) (see eTable 1). Tests for linear trend showed that SES was associated with health and well-being; those who reported a lower SES were more likely ($p<0.05$) to be unhealthy and to have a low level of well-being (Table 2).

Multiplicative interactions

Statistically significant ($p<0.05$) multiplicative interactions were observed between confounding variables, education, and indicators of SES (regressed on health and well-being). Among the respondents with 'very good' childhood SES, a higher education level was associated with slightly poorer health (eFigure 1). On the contrary, among respondents with 'very good', 'good' or 'difficult' childhood SES, a higher education level was associated with lower well-being (eFigure 2). Moreover, among the respondents with 'very good' childhood SES, a lower income (eFigure 3) and lower subjective social status (eFigure 4) was associated with slightly better health. Among the oldest group of respondents (age 65-74), lower occupational hierarchy position was associated with slightly better health (eFigure 5). Among the respondents with 'very low', 'low', or 'neither low nor high' subjective social status, a management position was associated with lower health (eFigure 6). Lastly, among the respondents with lowest levels of income or a low occupational hierarchy position, a higher management position was associated with lower well-being (eFigure 7, eFigure 8).

Education and health and well-being

Low education was associated with being unhealthy (RR_{TE}: 1.30, 95% CI: 1.27, 1.31), and having a low level of well-being (RR_{TE}: 1.13, 95% CI: 1.11, 1.15) (Table 3). Much of the association between education and health (RR_{NIE}: 1.23, 95% CI: 1.18, 1.25) and education and well-being (RR_{NIE}: 1.12, 95% CI: 1.10, 1.15) was mediated by all the mediators (Table 3). Occupational hierarchy position mediated most of the effect of education on health (RR_{NIE}: 1.14, 95% CI: 1.10, 1.17), while subjective social status mediated most of the effect of education on well-being (RR_{NIE}: 1.07, 95% CI: 1.07, 1.09) (Table 4).

Estimates calculated by both the traditional method for assessing mediation (difference approach) and the IOW method are presented in Table 3 and Table 4 for comparison purposes. Both approaches showed results that were in the same direction; however, the indirect effects were generally over-estimated with the traditional method (biased upwards). Consequently, generally the direct effects were under-estimated (biased downwards), as compared to IOW method. The direction and magnitude of bias in direct/indirect effects estimated with traditional method would depend on the magnitude and direction of interactions between the exposure, mediator(s), and covariates.

DISCUSSION

This article contributes to the literature on mediation analysis, as well as to the literature on the importance of education for health-related quality of life and subjective well-being. Theoretically, these findings are in line with the social causation hypothesis [77, 78], whereby social conditions, via lower education level, are associated with lower health and well-being in adulthood. The results indicate that most of the effect of education on health, and almost all of the effect of education on well-being, is mediated by SES.

Consistent with previous findings, SES mediated the effect of education on health and well-being [39-49]. However, this is the first study to present the influence of education on health and well-being through four different indicators of SES. Our findings have several implications for future research. First, they support the hypothesized mechanism that education is transformed into material benefits (in terms of income), as well as the perception of one's own social standing, which in turn translates into health and well-being. Second, this study contributes to the recognition that each indicator of SES may reflect a different dimension [79]. Third, it highlights subjective social status as part of a hypothesized mechanism through which education influences health and well-being. Fourth, occupational hierarchy position was closely in line with education and explained most of the association between education and health. However, it is plausible that different mechanisms apply to different participants. For instance, among a subset of our study sample, higher education may be associated with lower health and well-being due to unmet expectations and unrealised plans [80]. Other evidence has indicated that achieving a lower level of education than expected is associated with lower health and well-being in adulthood [81].

Education level (relative) tends to be stable over the generations, with children of parents with a low education level becoming adults with a low education level more often than expected by chance [14]. The social selection hypothesis assumes that health and well-being is a function of inter-generational [82] and intra-generational selection processes, whereby the healthy and able tend to acquire a higher education level. Indeed, intellectual functioning is, at least in part, based on genetic factors [83-85]. In this way, education level may represent a marker of socioeconomic advantage or disadvantage from one generation to the next. Some evidence [86], though not all [82, 84, 85, 87, 88], suggests that genetic factors do not explain observed inequalities in health by SES. Other evidence suggests that intelligence and cognitive ability in childhood are associated with health in adulthood independent of adult SES [89, 90].

Consistent with previous studies [14], significant interactions were observed between childhood SES and education. Previous studies examining the pathways between SES, and health and well-being, have been based on samples that were predominantly comprised of women (Finland [39]), predominantly comprised of men (Germany [40], U.S.A [46]), had a lower proportion of respondents with higher education (Finland [39], Sweden [40, 41], Germany [40], Canada [47], U.S.A [48]), a higher proportion of 'married' respondents (England [43]), a higher proportion of respondents with 'high' subjective social status (Japan [49]), and those respondents who were relatively older at baseline (England [43], U.S.A [46], Canada [47]). A key difference between Norway and most other countries is that education is entirely free at all levels. This may partly explain the high proportion of respondents with higher education (college or university) in our sample.

Some limitations should be acknowledged: the temporality between mediators and health and well-being cannot be determined empirically in this study. There may be some reverse causation, as those who had low SES might have been unhealthy. Therefore, the present study cannot determine whether SES was the cause, or the consequence of poor health or low well-being. Moreover, the possibility of mood congruency bias and differential measurement error cannot be ruled out [11]. The associations between education, SES, and health and well-being are not necessarily causative and deterministic; it is more likely that they are inter-related and probabilistic. Moreover, there may be unmeasured intermediate confounders [11-14]. For example, there may be other pathways (such as behavioral factors) that mediate the association between education and health [13, 14]. Several studies have shown that education is closely associated with socially patterned differences in factors such as alcohol intake and smoking [13, 91-93]. Other studies have consistently shown that these behavioural factors are associated with health and well-being in adulthood [13, 15]. Therefore, in the absence of alcohol intake and smoking in the models, the indirect effects presented here are likely to be over-estimated [13, 14]. On the other hand, non-differential measurement error in the mediators would lead to underestimated indirect effects (biased downwards) [13]. Consequently, the direct effects would be biased upwards [13].

The cut-offs used in this paper are aimed to separate the "perfectly healthy/highest well-being" from the rest. We performed the analyses using an alternative cut-off ("most unhealthy/lowest well-being" vs the rest) and obtained consistent results (in the same direction, see eTable 2 and eTable 3). Some of the mediators may confound the association between other mediators and health and well-being. Indeed, it is necessary to assess whether the indirect effect is due to the mediator, or some other variable that is associated with exposure, mediator, and outcome in the same direction [13]. Therefore, we also assessed the *independent* indirect effect via each mediator, adjusted for the remaining mediators (see eTable 4). Assessing the *independent* indirect effect of each mediator separately showed that the unique effect of education on health is mediated by income, management position, occupational hierarchy position, and subjective social status (eTable 4). However, for well-being, only income and subjective social status mediated the influence of education (eTable 4).

Despite the strength of the IOW approach in handling exposure-mediator, mediator-mediator, and mediator-confounder multiplicative interactions, there is one shortcomings of this approach. In order to estimate separate natural indirect effects in the context of multiple mediators (for example, two mediators: M1 and M2), one needs to assume that M1 is not an intermediate confounder in the exposure→M2→outcome association. For instance, it is assumed that management position and occupational hierarchy position do not temporally and causally affect subjective social status. This assumption may not be correct.

In conclusion, after adjusting for several confounding variables, the findings suggest that education is associated with health and well-being in adulthood and that most of this effect is mediated by indicators of SES. Mediation was assessed by the IOW approach, which does not rely on the ‘no interaction’ assumptions. We provide Stata code that makes it easy to assess mediation by the IOW approach in any MI dataset.

REFERENCES

1 Robins JM, Greenland S. Identifiability and Exchangeability for Direct and Indirect Effects. *Epidemiology* 1992;**3**:143-55.

2 Hayes AF. *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*: Guilford Press 2013.

3 Preacher KJ. Advances in Mediation Analysis: A Survey and Synthesis of New Developments. *Annual Review of Psychology* 2015;**66**:825-52.

4 MacKinnon DP. *Introduction to Statistical Mediation Analysis*. New York: Taylor & Francis 2008.

5 Lange T, Rasmussen M, Thygesen LC. Assessing Natural Direct and Indirect Effects Through Multiple Pathways. *American Journal of Epidemiology* 2014;**179**:513-8.

6 Hong G. *Causality in a Social World: Moderation, Meditation and Spill-over*: John Wiley & Sons, Ltd 2015.

7 Petersen ML, Sinisi SE, van der Laan MJ. Estimation of direct causal effects. *Epidemiology* 2006;**17**:276-84.

8 Thomas RTH, Joffe MM, Lynch KG*, et al.* Causal Mediation Analyses with Rank Preserving Models. *Biometrics* 2007;**63**:926-34.

9 Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol* 1986;**51**:1173-82.

10 Judd CM, Kenny DA. Process Analysis: Estimating Mediation in Treatment Evaluations. *Evaluation Review* 1981;**5**:602-19.

11 Sheikh MA, Abelsen B, Olsen JA. Differential recall bias, intermediate confounding, and mediation analysis in life course epidemiology: An analytic framework with empirical example. *Frontiers in Psychology* 2016;**7**.

12 Sheikh MA. Childhood adversities and chronic conditions: Examination of mediators, recall bias and age at diagnosis. *International Journal of Public Health* 2017.

13 Sheikh MA. Confounding and statistical significance of indirect effects: Childhood adversity, education, smoking, and anxious and depressive symptomatology. *Frontiers in Psychology* 2017;**8**.

14 Sheikh MA, Abelsen B, Olsen JA. Role of respondents' education as a mediator and moderator in the association between childhood socio-economic status and later health and wellbeing. *BMC Public Health* 2014;**14**:1172.

15 Sheikh MA, Abelsen B, Olsen JA. Clarifying associations between childhood adversity, social support, behavioral factors, and mental health, health, and well-being in adulthood: A population-based study. *Frontiers in Psychology* 2016;**7**.

16 Lamadrid-Figueroa H, Téllez-Rojo MM, Ángeles G*, et al.* Bias correction by use of errors-in-variables regression models in studies with K-X ray fluorescence bone lead measurements. *Environmental research* 2011;**111**:17-20.

17 Cai L. Potential Applications of Latent Variable Modeling for the Psychometrics of Medical Simulation. *Military medicine* 2013;**178**:115-20.

18 Schofield LS. CORRECTING FOR MEASUREMENT ERROR IN LATENT VARIABLES USED AS PREDICTORS(). *The annals of applied statistics* 2015;**9**:2133-52.

19 Fergusson DM, Horwood LJ, Boden JM. Structural equation modeling of repeated retrospective reports of childhood maltreatment. *International Journal of Methods in Psychiatric Research* 2011;**20**:93-104.

20 Imai K, Keele L, Tingley D. A General Approach to Causal Mediation Analysis. *Psychol Methods* 2010;**15**:309 –34.

21 Hicks R, Tingley D. Causal mediation analysis. *The Stata Journal* 2011;**11**:605-19.

22 Robins J. The control of confounding by intermediate variables. *Stat Med* 1989;**8**:679-701.

23 Robins J. A new approach to causal inference in mortality studies with a sustained exposure period--Application to control of the healthy worker survivor effect. *Mathematical Modelling* 1986;**7**:1393-512.

24 Robins JM. The analysis of randomized and nonrandomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In: Sechrest L, Freeman H, Mulley A, eds. *Health Service*

Research Methodology: A Focus on AIDS. Washington, DC: U.S. Public Health Service, National Center for Health Services Research 1989.

25 Robins JM, Blevins D, Ritter G*, et al.* G-Estimation of the Effect of Prophylaxis Therapy for Pneumocystis carinii Pneumonia on the Survival of AIDS Patients. *Epidemiology* 1992;**3**:319-36.

26 Robins JM, Hernán MÁ, Brumback B. Marginal Structural Models and Causal Inference in Epidemiology. *Epidemiology* 2000;**11**:550-60.

27 Fryers T, Melzer D, Jenkins R. Social inequalities and the common mental disorders. *Soc Psychiatry Psychiatr Epidemiol* 2003;**38**:229-37.

28 Ross CE, Wu C-l. The Links Between Education and Health. *American Sociological Review* 1995;**60**:719-45.

29 Elder GHJ. Perspectives on the Life Course. In: Elder GHJ, Council SSR, eds. *Life course dynamics: trajectories and transitions, 1968-1980*. New York: Cornell University Press 1985.

30 Pearlin LI. The sociological study of stress. *J Health Soc Behav* 1989;**30**:241-56.

31 Turner RJ, Wheaton B, Lloyd DA. The Epidemiology of Social Stress. *American Sociological Review* 1995;**60**:104-25.

32 Lynch JW, Kaplan GA, Shema SJ. Cumulative Impact of Sustained Economic Hardship on Physical, Cognitive, Psychological, and Social Functioning. *New England Journal of Medicine* 1997;**337**:1889-95.

33 Mackie KS, Holahan CK, Gottlieb NH. Employee Involvement Management Practices, Work Stress, and Depression in Employees of a Human Services Residential Care Facility. *Human Relations* 2001;**54**:1065-92.

34 Wiesner M, Windle M, Freeman A. Work stress, substance use, and depression among young adult workers: an examination of main and moderator effect model. *Journal of occupational health psychology* 2005;**10**:83-96.

35 Zimmerman FJ, Christakis DA, Vander Stoep A. Tinker, tailor, soldier, patient: work attributes and depression disparities among young adults. *Social Science & Medicine* 2004;**58**:1889-901.

36 Baum A, Garofalo JP, Yali AM. Socioeconomic Status and Chronic Stress: Does Stress Account for SES Effects on Health? *Annals of the New York Academy of Sciences* 1999;**896**:131-44.

37 McFarland MJ, Wagner B, Marklin S. College Education and Sense of Control. *Socius* 2016;**2**:2378023116656011.

38 Edgerton JD, Roberts LW, von Below S. Education and Quality of Life. In: Land KC, Michalos AC, Sirgy MJ, eds. *Handbook of Social Indicators and Quality of Life Research*. Dordrecht: Springer Netherlands 2012:265-96.

39 Lahelma E, Martikainen P, Laaksonen M*, et al.* Pathways between socioeconomic determinants of health. *Journal of Epidemiology and Community Health* 2004;**58**:327-32.

40 Geyer S, Hemström Ö, Peter R*, et al.* Education, income, and occupational class cannot be used interchangeably in social epidemiology. Empirical evidence against a common practice. *Journal of Epidemiology and Community Health* 2006;**60**:804-10.

41 Torssander J, Erikson R. Stratification and Mortality—A Comparison of Education, Class, Status, and Income. *European Sociological Review* 2010;**26**:465-74.

42 Gjonça E, Tabassum F, Breeze E. Socioeconomic differences in physical disability at older age. *Journal of Epidemiology and Community Health* 2009;**63**:928-35.

43 Demakakos P, Nazroo J, Breeze E*, et al.* Socioeconomic status and health: The role of subjective social status. *Social Science & Medicine* 2008;**67**:330-40.

44 Miyakawa M, Magnusson Hanson LL, Theorell T*, et al.* Subjective social status: its determinants and association with health in the Swedish working population (the SLOSH study). *The European Journal of Public Health* 2012;**22**:593-7.

45 Desjardins R, Schuller T. Measuring the effects of education on health and civic engagement: proceedings of the copenhagen symposium. Copenhagen: OECD 2006.

46 Zhang W, Chen Q, McCubbin H*, et al.* Predictors of mental and physical health: Individual and neighborhood levels of education, social well-being, and ethnicity. *Health & Place* 2011;**17**:238-47.

47 White CM, St. John PD, Cheverie MR*, et al.* The role of income and occupation in the association of education with healthy aging: results from a population-based, prospective cohort study. *BMC Public Health* 2015;**15**:1-11.

48 Dennis EF, Webb DA, Lorch SA*, et al.* Subjective Social Status and Maternal Health in a Low Income Urban Population. *Maternal and Child Health Journal* 2011;**16**:834-43.

49 Honjo K, Kawakami N, Tsuchiya M*, et al.* Association of Subjective and Objective Socioeconomic Status with Subjective Mental Health and Mental Disorders Among Japanese Men and Women. *International Journal of Behavioral Medicine* 2013;**21**:421-9.

50 Kaufman JS, Maclehose RF, Kaufman S. A further critique of the analytic strategy of adjusting for covariates to identify biologic mediation. *Epidemiol Perspect Innov* 2004;**1**:4.

51 Clogg CC, Petkova E, Shihadeh ES. Statistical Methods for Analyzing Collapsibility in Regression Models. *Journal of Educational and Behavioral Statistics* 1992;**17**:51-74.

52 Preacher KJ, Hayes AF. SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments & Computers* 2004;**36**:717-31.

53 Robins JM. Semantics of causal dag models and the identification of direct and indirect effects. In: Green P, Hjort N, Richardson S, eds. *Highly Structured Stochastic Systems*. New York: Oxford University Press 2003:70-81.

54 VanderWeele T, Vansteelandt S. Mediation Analysis with Multiple Mediators. *Epidemiologic Methods* 2014:95.

55 Emsley R, Liu H. PARAMED: Stata module to perform causal mediation analysis using parametric regression models. Statistical Software Components S457581, Boston College Department of Economics 2013.

56 Valeri L, Vanderweele TJ. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: Theoretical assumptions and implementation with SAS and SPSS macros. *Psychol Methods* 2013;**18**:137-50.

57 Imai K, Keele L, Tingley D*, et al.* Causal Mediation Analysis Using R. In: Vinod DH, ed. *Advances in Social Science Research Using R*. New York, NY: Springer New York 2010:129-54.

58 Imai K, Keele L, Yamamoto T. Identification, Inference and Sensitivity Analysis for Causal Mediation Effects. 2010:51-71.

59 Etherington N. Re-evaluating gender differences in self-rated health: The importance of cohort. *Journal of Women & Aging* 2017;**29**:150-62.

60 Goosby BJ. Early Life Course Pathways of Adult Depression and Chronic Pain. *Journal of Health and Social Behavior* 2013;**54**:75-91.

61 Schaan B. The interaction of family background and personal education on depressive symptoms in later life. *Social Science & Medicine* 2014;**102**:94-102.

62 Daniel RM, De Stavola BL, Cousens SN*, et al.* Causal mediation analysis with multiple mediators. *Biometrics* 2015;**71**:1-14.

63 Tchetgen Tchetgen EJ. Inverse odds ratio-weighted estimation for causal mediation analysis. *Statistics in Medicine* 2013;**32**:4567-80.

64 Nguyen QC, Osypuk TL, Schmidt NM*, et al.* Practical Guidance for Conducting Mediation Analysis With Multiple Mediators Using Inverse Odds Ratio Weighting. *American Journal of Epidemiology* 2015;**181**:349-56.

65 Buis ML. Direct and indirect effects in a logit model. *Stata Journal* 2010;**10**:11-29.

66 Breen R, Karlson KB, Holm A. Total, Direct, and Indirect Effects in Logit and Probit Models. *Sociological Methods & Research* 2013;**42**:164-91.

67 Karlson KB, Holm A, Breen R. Comparing Regression Coefficients Between Same-sample Nested Models Using Logit and Probit. *Sociological Methodology* 2012;**42**:286-313.

68 Karlson KB, Holm A. Decomposing primary and secondary effects: A new decomposition method. *Research in Social Stratification and Mobility* 2011;**29**:221-37.

69 Emsley RA, Liu H, Dunn G*, et al.* Paramed: A command to perform causal mediation analysis using parametric models. *The Stata Journal (in preparation)* 2014.

70 Jann B. FAIRLIE: Stata module to generate nonlinear decomposition of binary outcome differentials. Statistical Software Components S456727, Boston College Department of Economics 2006.

71 Mehmetoglu M. MEDSEM: Stata module to perform mediation analysis using structural equation modelling. Statistical Software Components S458300, Boston College Department of Economics. 2017.

72 Deutsch J, Jacobus M, Vigil A. RMPW: Stata module to implement the RMPW method of causal mediation analysis to decompose treatment effects into "direct" and "indirect" effects. Statistical Software Components S458301, Boston College Department of Economics. 2017.

73 Mehmetoglu M. VAF: Stata module to compute mediation effect in SEM. Statistical Software Components S458119, Boston College Department of Economics. 2015.

74 Sinning M, Hahn M, Bauer TK. The Blinder-Oaxaca decomposition for nonlinear regression models. *Stata Journal* 2008;**8**:480-92.

75 Jacobsen BK, Eggen AE, Mathiesen EB*, et al.* Cohort profile: The Tromsø Study. *International Journal of Epidemiology* 2012;**41**:961-7.

76 Operario D, Adler NE, Williams DR. Subjective social status: reliability and predictive utility for global health. *Psychology & Health* 2004;**19**:237-46.

77 Johnson JG, Cohen P, Dohrenwend BP*, et al.* A longitudinal investigation of social causation and social selection processes involved in the association between socioeconomic status and psychiatric disorders. *Journal of abnormal psychology* 1999;**108**:490-9.

78 Mossakowski KN. Social Causation and Social Selection. *The Wiley Blackwell Encyclopedia of Health, Illness, Behavior, and Society*: John Wiley & Sons, Ltd 2014.

79 Camelo LdV, Giatti L, Barreto SM. Subjective social status, self-rated health and tobacco smoking: Brazilian Longitudinal Study of Adult Health (ELSA-Brasil). *Journal of Health Psychology* 2014;**19**:1388-99.

80 Reynolds JR, Baird CL. Is There a Downside to Shooting for the Stars? Unrealized Educational Expectations and Symptoms of Depression. *American Sociological Review* 2010;**75**:151-72.

81 Mossakowski KN. Unfulfilled expectations and symptoms of depression among young adults. *Social Science & Medicine* 2011;**73**:729-36.

82 Bombay A, Matheson K, Anisman H. The impact of stressors on second generation Indian residential school survivors. *Transcultural Psychiatry* 2011;**48**:367-91.

83 Masten AS, Coatsworth JD. The development of competence in favorable and unfavorable environments: Lessons from research on successful children. *American Psychologist* 1998;**53**:205-20.

84 Mayer SE. *What Money Can't Buy: Family Income and Children's Life Chances*: Harvard University Press 1997.

85 Marioni RE, Ritchie SJ, Joshi PK*, et al.* Genetic variants linked to education predict longevity. *Proceedings of the National Academy of Sciences* 2016;**113**:13366-71.

86 Adler NE, Boyce T, Chesney MA*, et al.* Socioeconomic status and health: The challenge of the gradient. *American Psychologist* 1994;**49**:15-24.

87 Stansfeld SA, Clark C, Rodgers B*, et al.* Repeated exposure to socioeconomic disadvantage and health selection as life course pathways to mid-life depressive and anxiety disorders. *Soc Psychiatry Psychiatr Epidemiol* 2011;**46**:549-58.

88 Elovainio M, Pulkki-Raback L, Jokela M*, et al.* Socioeconomic status and the development of depressive symptoms from childhood to adulthood: a longitudinal analysis across 27 years of follow-up in the Young Finns study. *Soc Sci Med* 2012;**74**:923-9.

89 Hart CL, Taylor MD, Smith GD*, et al.* Childhood IQ and cardiovascular disease in adulthood: prospective observational study linking the Scottish Mental Survey 1932 and the Midspan studies. *Social Science & Medicine* 2004;**59**:2131-8.

90 Hagger-Johnson G, Mõttus R, Craig LCA*, et al.* Pathways from childhood intelligence and socioeconomic status to late-life cardiovascular disease risk. *Health Psychology* 2012;**31**:403-12.

91 Jorm AF, Rodgers B, Jacomb PA*, et al.* Smoking and mental health: results from a community survey. *Med J Aust* 1999;**170**:74-7.

92 Gilman SE, Martin LT, Abrams DB*, et al.* Educational attainment and cigarette smoking: a causal association?†. *International Journal of Epidemiology* 2008;**37**:615-24.

93 Assari S, Lankarani MM. Education and Alcohol Consumption among Older Americans; Black–White Differences. *Frontiers in Public Health* 2016;**4**.

Table 1. General Characteristics of the Study Sample in the Un-imputed dataset and the dataset with Multiple Imputations (n=8,516)

Characteristics		Un-imputed dataset		Dataset with multiple imputations
		n[a]	%	%
Age (years)[b]	25-34	1656	19.4	- [b]
	35-44	1741	20.5	- [b]
	45-54	3138	36.9	- [b]
	55-64	1545	18.1	- [b]
	65-74	436	5.1	- [b]
Gender[b]	Male	4008	47.1	- [b]
	Female	4508	52.9	- [b]
Mother's psychological problems[b]	No	7959	93.5	- [b]
	Yes	557	6.5	- [b]
Father's psychological problems[b]	No	8284	97.3	- [b]
	Yes	232	2.7	- [b]
Exposure to passive smoke in childhood	No	2193	25.8	25.8
	Yes	6318	74.2	74.2
Marital Status	Single	1766	20.8	20.8
	Married or registered partnership	5526	65.0	64.9
	Widowed, divorced, or separated	1207	14.2	14.2
Childhood socioeconomic status	Very good or good	5278	66.8	66.8
	Difficult or very difficult	2626	33.2	33.2
Having enough friends	No	1434	18.3	18.4
	Yes	6414	81.7	81.6
Number of friends	0-3	3112	42.2	42.2
	4-5	2098	28.4	28.4
	6 or more	2169	29.4	29.4
Physical activity (hours/week)	Less than 1	2063	24.3	24.4
	1-2	3091	36.5	36.5
	3 or more	3324	39.2	39.2
Education	Low	5829	68.6	68.6
	High	2663	31.4	31.4
Income	Low (up to 400,000NOK)	3039	38.3	40.1
	Middle (401,000-700,000NOK)	2947	37.2	36.5
	High (701,000NOK and above)	1944	24.5	23.4
Management position	No	4396	64.4	65.4
	Low	1636	24.0	23.3
	High	795	11.6	11.2
Occupational hierarchy position	Low	1923	26.7	29.5
	Middle	3719	51.5	50.9
	High	1575	21.8	19.6
Subjective social status	Low	2722	35.2	34.6
	Neither low nor high	4419	57.2	57.5
	High	587	7.6	7.9
Health[c]	Unhealthy	4291	55.0	55.0
	Healthy	3511	45.0	45.0
Well-being[d]	Low	5804	68.2	68.2
	High	2712	31.9	31.8

[a] The numbers do not add up to 8,516 due to missing values.

[b] There were no missing values, so no imputations were made for these variables.

[c] Health was assessed by the EQ-5D generic measure of health-related quality of life

[d] Well-being was measured by the satisfaction with life scale (SWLS)

Table 2. Distribution (%) of Adult Socioeconomic Status (SES) and Health and Well-being by level of Education, and SES by Health and Well-being (n=8,516)

		Education			Health[a]			Well-being[b]		
		Low n=5,829	High n=2,663	Test statistic	Unhealthy n=4,291	Healthy n=3,511	Test statistic	Low 5,804	High 2,712	Test statistic
		%	%		%	%		%	%	
Income[c,e]	Low	48.4	17.5	χ^2 (2) = 1176 [f]	44.0	33.3	χ^2 (2)=268.36 [f]	43.4	27.7	χ^2 (2)=242.77 [f]
	Middle	37.7	36.3		37.6	37.7		36.4	38.7	
	High	14.0	46.3		18.4	29.0		20.2	33.6	
Management position[e]	No	70.4	52.7	χ^2 (2) = 309.70 [f]	68.8	58.8	χ^2 (2)=81.86 [f]	66.2	60.9	χ^2 (2)= 27.03 [f]
	Low	22.4	27.1		21.9	26.7		23.5	25.0	
	High	7.2	20.3		9.3	14.6		10.4	14.2	
Occupational hierarchy[e]	Low	39.6	2.6	χ^2 (2) = 2455 [f]	30.6	21.5	χ^2 (2)=166.08 [f]	29.9	21.8	χ^2 (2)=70.39 [f]
	Middle	54.7	45.7		52.6	50.1		51.5	51.6	
	High	5.7	51.7		16.8	28.4		19.4	26.6	
Subjective social status[d,e]	Low	10.2	2.1	χ^2 (2) = 857.86 [f]	9.4	5.3	χ^2 (2)=150.38 [f]	9.1	4.8	χ^2 (2)=214.33 [f]
	Similar	65.4	40.2		60.8	52.5		61.3	49.3	
	High	24.4	57.7		29.8	42.2		29.7	45.9	
Health[a,e]	Unhealthy	60.2	44.0	χ^2 (1) = 179.86 [f]						
	Healthy	39.8	55.9							
Well-being[b,e]	Low	71.0	61.6	χ^2 (1) = 74.64 [f]						
	High	29.0	38.4							

[a]EQ-5D: Health was assessed by the EQ-5D generic measure of health-related quality of life
[b]SWLS: Well-being was measured by the satisfaction with life scale (SWLS)
[c]Income: Low (up to 400,000NOK), Middle (401,000-700,000NOK), High (701,000NOK and above)
[d]Subjective social status: Similar (Neither low nor high)
[e]Test for linear trend $p<0.05$
[f] $p<0.05$

Table 3: Total, Direct and Indirect Effect of Education on Health and Well-being (n=8,516)

		Health[a]					
		Total effect		Natural direct effect		Natural indirect effect	
		RR	95% CI	RR	95% CI	RR	95% CI
Low education[c]	Traditional	1.30	1.27, 1.31	1.08	1.03, 1.15	1.20	1.15, 1.24
	IOW	1.30	1.27, 1.31	1.06	1.02, 1.10	1.23	1.18, 1.25
		Well-being[b]					
Low education[c]	Traditional	1.13	1.11, 1.15	0.99	0.96, 1.00	1.14	1.12, 1.15
	IOW	1.13	1.11, 1.15	1.00	0.99, 1.02	1.12	1.10, 1.15

[a]EQ-5D: Health was assessed by the EQ-5D generic measure of health-related quality of life. Healthy for EQ-5D included all respondents ticking level one for all five dimensions
[b]SWLS: Well-being was measured by the first three items from the satisfaction with life scale measured on a 7-point scale. High wellbeing included those who reported 6 or 7 for all three items..
[c]Education: (Low= primary and secondary school or similar (i.e. 7–10 years of schooling) or vocational school or high school,) (High (reference)= college or university (less than 4 years) or college or university (4 years or more))
All models adjusted for age, gender, exposure to passive smoke in childhood, having enough friends, number of friends, marital status, physical activity, mothers' psychological problems, fathers' psychological problems, and childhood socioeconomic status.

Table 4: Total, Direct and Indirect Effect of Education on Health and Well-being (n=8,516)

		Health[a]						Well-being[b]					
		Total effect		Direct effect		Indirect effect		Total effect		Direct effect		Indirect effect	
		RR	95% CI	RR	95% CI	RR	95% CI	RR	95% CI	RR	95% CI	RR	95% CI
		Only income											
Low education[c]	Traditional	1.30	1.27, 1.31	1.19	1.17, 1.24	1.09	1.07, 1.11	1.13	1.11, 1.15	1.04	1.01, 1.10	1.08	1.07, 1.09
	IOW	1.30	1.27, 1.31	1.22	1.20, 1.27	1.06	1.04, 1.10	1.13	1.11, 1.15	1.06	1.02, 1.13	1.06	1.05, 1.08
		Only management position											
Low education[c]	Traditional	1.30	1.27, 1.31	1.27	1.24, 1.32	1.02	1.01, 1.03	1.13	1.11, 1.15	1.11	1.09, 1.13	1.01	1.01, 1.02
	IOW	1.30	1.27, 1.31	1.28	1.25, 1.33	1.01	1.01, 1.02	1.13	1.11, 1.15	1.12	1.10, 1.13	1.01	1.00, 1.01
		Only occupational hierarchy position											
Low education[c]	Traditional	1.30	1.27, 1.31	1.15	1.10, 1.19	1.13	1.10, 1.16	1.13	1.11, 1.15	1.06	1.04, 1.10	1.06	1.04, 1.07
	IOW	1.30	1.27, 1.31	1.14	1.10, 1.21	1.14	1.10, 1.17	1.13	1.11, 1.15	1.07	1.06, 1.18	1.05	1.01, 1.07
		Only subjective social status											
Low education[c]	Traditional	1.30	1.27, 1.31	1.23	1.20, 1.27	1.06	1.04, 1.06	1.13	1.11, 1.15	1.05	1.02, 1.06	1.08	1.06, 1.08
	IOW	1.30	1.27, 1.31	1.25	1.22, 1.31	1.04	1.01, 1.05	1.13	1.11, 1.15	1.05	1.03, 1.06	1.07	1.07, 1.09

[a]EQ-5D: Health was assessed by the EQ-5D generic measure of health-related quality of life. Healthy for EQ-5D included all respondents ticking level one for all five dimensions

[b]SWLS: Well-being was measured by the first three items from the satisfaction with life scale measured on a 7-point scale. High wellbeing included those who reported 6 or 7 for all three items..

[c]Education: (Low= primary and secondary school or similar (i.e. 7–10 years of schooling) or vocational school or high school,) (High (reference)= college or university (less than 4 years) or college or university (4 years or more))

All models adjusted for potential confounding variables (age, gender, exposure to passive smoke in childhood, having enough friends, number of friends, marital status, physical activity, mothers' psychological problems, fathers' psychological problems, and childhood socioeconomic status.)

Supplementary Digital Content

Citation:

Sheikh, M. A., Abelsen, B., & Olsen, J. A. (2017). Education and health and well-being: direct and indirect effects with multiple mediators and interactions with multiple imputed data in Stata. *Journal of Epidemiology and Community Health, 71*(11), 1037-1045.

doi:10.1136/jech-2016-208671

eAppendix 1:Stata Code for assessing mediation with a group of mediators (M1, M2, M3, M4) together.

Sentences in green starting with (*) are not part of the computer code, and serve only as an explanation of the process. The Stata code is coloured in blue. The highlighted parts are names of the variables, which can be replaced accordingly.

```
*For all mediators (M1 M2 M3 M4) together.

*Code for estimating natural direct, natural indirect, and total effect using inverse odds weighting with multiple imputed data (MI) in Stata

*The Stata code provided given below can be followed for any multiple imputed dataset in Stata, by replacing
*the exposure, mediator (s), outcome, and confounding variables accordingly.

program IOWMI , rclass

*Retain estimates of predicted probability, inverse odds, and inverse odds weights for later use ;

capture drop linpred predprob inverseodds wt_iow

* Fit a logistic regression model for education (0=high, 1=low) conditional on the mediators and

*confounding variables.

mi estimate, saving(miest):  logit education M1 M2 M3 M4 confounding_variables

*Calculate linear prediction for each observation and use that to calculate predicted probabilities and inverse odds;

mi predict linpred using miest, xb

mi passive: gen predprob = exp(linpred)/(1+exp(linpred))

mi passive: gen inverseodds = ((1-predprob)/predprob)

*Calculate inverse odds weights;

*Assign the IOW of each observation in the unexposed group (education= 0 (high)) equal to 1.

mi passive: gen wt_iow = 1 if education==0

* Compute an IOW by taking the inverse of the predicted log odds for each observation in the

*exposed group (education=1 (low)).

mi passive: replace wt_iow = inverseodds if education==1
```

```
* Estimate the total effect of education using a generalized linear model (family=Poisson) of the

*regression of the outcome on education and confounding variables, with link=log function;

mi estimate, saving(miest1) eform post: glm outcome education confounding_variables  , fam(poisson) link(log) vce(robust)

 matrix bb_total= e(b_mi)

scalar b_total=(bb_total[1,1])

return scalar b_total=bb_total[1,1]


* Estimate the natural direct effect of education via weighted generalized linear model *(family=Poisson) of the regression of the outcome on education and confounding factors, with *link=log function and the weights obtained earlier

mi estimate, saving(miest2) eform post: glm outcome education confounding_variables [pweight=wt_iow], fam(poisson) link(log) vce(robust)

matrix bb_direct = e(b_mi)

scalar b_direct=(bb_direct[1,1])

return scalar b_direct=bb_direct[1,1]

* Calculate the natural indirect effects of education on the outcome via the proposed mediators by subtracting the direct effects from the total effects as;

return scalar b_indirect = b_total-b_direct

end

*Estimate 95% confidence intervals with bootstrapping

bootstrap exp(r(b_indirect)) exp(r(b_direct)) exp(r(b_total)), seed(12345) reps(100): IOWMI

estat bootstrap, all
```

eAppendix 2: Stata Code for assessing mediation with a single mediator (M1) separately (assuming that other mediators are not confounders):

Sentences in green starting with (*) are not part of the computer code, and serve only as an explanation of the process. The Stata code is coloured in blue. The highlighted parts are names of the variables, which can be replaced accordingly.

*Code for estimating natural direct, natural indirect, and total effect using inverse odds weighting with multiple imputed data (MI) in Stata

*The Stata code provided given below can be followed for any multiple imputed dataset in Stata, by replacing
*the exposure, mediator, outcome, and confounding variables accordingly.

```
program IOWMI , rclass
```

*Retain estimates of predicted probability, inverse odds, and inverse odds weights for later use ;

```
capture drop linpred predprob inverseodds wt_iow
```

* Fit a logistic regression model for education (0=high, 1=low) conditional on the mediator and

*confounding variables.

```
mi estimate, saving(miest):  logit education M1 confounding_variables
```

*Calculate linear prediction for each observation and use that to calculate predicted probabilities and inverse odds;

```
mi predict linpred using miest, xb

mi passive: gen predprob = exp(linpred)/(1+exp(linpred))

mi passive: gen inverseodds = ((1-predprob)/predprob)
```

*Calculate inverse odds weights;

*Assign the IOW of each observation in the unexposed group (education= 0 (high)) equal to 1.

```
mi passive: gen wt_iow = 1 if education==0
```

* Compute an IOW by taking the inverse of the predicted log odds for each observation in the

*exposed group (education=1 (low)).

```
mi passive: replace wt_iow = inverseodds if education==1
```

* Estimate the total effect of education using a generalized linear model (family=Poisson) of the

*regression of the outcome on education and confounding variables, with link=log function;

```
mi estimate, saving(miest1) eform post: glm outcome education confounding_variables  , fam(poisson) link(log) vce(robust)

matrix bb_total= e(b_mi)

scalar b_total=(bb_total[1,1])

return scalar b_total=bb_total[1,1]
```

* Estimate the natural direct effect of education via weighted generalized linear model *(family=Poisson) of the regression of the outcome on education and confounding factors, with link=log function and the weights obtained earlier

```
mi estimate, saving(miest2) eform post: glm outcome education confounding_variables [pweight=wt_iow], fam(poisson) link(log) vce(robust)

matrix bb_direct = e(b_mi)

scalar b_direct=(bb_direct[1,1])

return scalar b_direct=bb_direct[1,1]

* Calculate the natural indirect effects of education on the outcome via the proposed mediator by subtracting the direct effects from the total effects as;

return scalar b_indirect = b_total-b_direct

end

*Estimate 95% confidence intervals with bootstrapping

*Note that only 100 replications are requested below, and number can be increased accordingly

bootstrap exp(r(b_indirect)) exp(r(b_direct)) exp(r(b_total)), seed(12345) reps(100): IOWMI

estat bootstrap, all
```

eAppendix 3: Stata Code for assessing mediation with a single mediator (M1) separately (adjusted for other mediators):

Sentences in green starting with (*) are not part of the computer code, and serve only as an explanation of the process. The Stata code is coloured in blue. The highlighted parts are names of the variables, which can be replaced accordingly.

*Code for estimating natural direct, natural indirect, and total effect using inverse odds weighting with multiple imputed data (MI) in Stata

*The Stata code provided given below can be followed for any multiple imputed dataset in Stata, by replacing *the exposure, mediator (s), outcome, and confounding variables accordingly.

```
program IOWMI , rclass
```

*Retain estimates of predicted probability, inverse odds, and inverse odds weights for later use ;

```
capture drop linpred predprob inverseodds wt_iow
```

* Fit a logistic regression model for education (0=high, 1=low) conditional on the mediators and

*confounding variables.

*Note that the mediator of interest is M1, while the remaining mediators (M2 M3 M4) serve as confounding

variables in the analysis.

*Note that it should be assumed that the mediators M2, M3, and M4 are not intermediate confounders in the *model

*Education→M1→Outcome

```
mi estimate, saving(miest):  logit education M1 M2 M3 M4 confounding_variables
```

*Calculate linear prediction for each observation and use that to calculate predicted probabilities and inverse odds;

```
mi predict linpred using miest, xb

mi passive: gen predprob = exp(linpred)/(1+exp(linpred))

mi passive: gen inverseodds = ((1-predprob)/predprob)
```

*Calculate inverse odds weights;

*Assign the IOW of each observation in the unexposed group (education= 0 (high)) equal to 1.

```
mi passive: gen wt_iow = 1 if education==0
```

* Compute an IOW by taking the inverse of the predicted log odds for each observation in the

*exposed group (education=1 (low)).

```
mi passive: replace wt_iow = inverseodds if education==1
```

* Estimate the total effect of education using a generalized linear model (family=Poisson) of the

*regression of the outcome on education and confounding variables, with link=log function;

```
mi estimate, saving(miest1) eform post: glm outcome education M2 M3 M4 confounding_variables ,
fam(poisson) link(log) vce(robust)

matrix bb_total= e(b_mi)
```

```
scalar b_total=(bb_total[1,1])

return scalar b_total=bb_total[1,1]

* Estimate the natural direct effect of education via weighted generalized linear model *(family=Poisson) of the regression of the outcome on education and confounding factors, with *link=log function and the weights obtained earlier

mi estimate, saving(miest2) eform post: glm outcome education M2 M3 M4 confounding_variables [pweight=wt_iow], fam(poisson) link(log) vce(robust)

matrix bb_direct = e(b_mi)

scalar b_direct=(bb_direct[1,1])

return scalar b_direct=bb_direct[1,1]

* Calculate the natural indirect effects of education on the outcome via the proposed mediator by subtracting the direct effects from the total effects as;

return scalar b_indirect = b_total-b_direct

end

*Estimate 95% confidence intervals with bootstrapping

*Note that only 100 replications are requested below, and number can be increased accordingly

bootstrap exp(r(b_indirect)) exp(r(b_direct)) exp(r(b_total)), seed(12345) reps(100): IOWMI

estat bootstrap, all
```

eTable 1: Pearson Correlation Between Indicators of Adult Socioeconomic Status (SES).

	Income		Management position		Occupational hierarchy position		Subjective social status
	r	95% CI	r	95% CI	r	95% CI	r
Income	1.00						
Management position	0.22	0.20, 0.24	1.00				
Occupational hierarchy position	0.43	0.41, 0.44	0.26	0.24, 0.28	1.00		
Subjective social status	0.22	0.20, 0.24	0.28	0.26, 0.30	0.34	0.32, 0.36	1.00

eTable 2. Proportion (%) of Health and Well-being in the Un-imputed Dataset, and in the Imputed Dataset with Multiple Imputation with Alternative Cut offs (n=8,516)

		Un-imputed dataset	Imputed dataset
		%	%
Health and well-being			
- Health (EQ-5D)	Healthy	89.2	88.7
	Unhealthy	10.8	11.3
- Well-being (SWLS)	High	99.0	98.1
	Low	0.9	1.9

EQ-5D: Health was assessed by the EQ-5D generic measure of health-related quality of life. The sum of five indicators (range: 5–15) was divided in two groups. Those with the scores 5-7 were considered as healthy, while those with the scores 8 and above were considered as unhealthy.
SWLS: Well-being was measured by the first three items from the satisfaction with life scale measured on a 7-point scale. The sum of three indicators (range: 3-21) was divided in two groups. Those with the scores 6-21 were considered as having a high level of well-being, while those with the scores 3-5 were considered as having a low level of well-being

eTable 3: Total, Natural Direct and Natural Indirect Effect of Education on Health and Well-being with Alternative Cut offs (n=8,516)

		Health[a]					
		Total effect		Natural direct effect		Natural indirect effect	
		RR	95% CI	RR	95% CI	RR	95% CI
Education[c]	High	1.00		1.00		1.00	
	Low	1.65	1.45, 1.83	1.20	0.99, 1.39	1.37	1.18, 1.56
		Well-being[b]					
Education[c]	High	1.00		1.00		1.00	
	Low	1.83	1.63, 2.12	1.17	0.54, 1.65	1.55	1.23, 2.35

[a]EQ-5D: Health was assessed by the EQ-5D generic measure of health-related quality of life. The sum of five indicators (range: 5–15) was divided in two groups. Those with the scores 5-7 were considered as healthy, while those with the scores 8 and above were considered as unhealthy.
[b]SWLS: Well-being was measured by the first three items from the satisfaction with life scale measured on a 7-point scale. The sum of three indicators (range: 3-21) was divided in two groups. Those with the scores 6-21 were considered as having a high level of well-being, while those with the scores 3-5 were considered as having a low level of well-being

[c]Education: (Low= primary and secondary school or similar (i.e. 7–10 years of schooling) or vocational school or high school,) (High= college or university (less than 4 years) or college or university (4 years or more))

All models adjusted for age, gender, exposure to passive smoke in childhood, having enough friends, number of friends, marital status, physical activity, mothers' psychological problems, fathers' psychological problems, and childhood socioeconomic status.

eTable 4: Total, Natural Direct and Natural Indirect Effect of Education on Health and Well-being (mediator-adjusted) (n=8,516)

		Health[a]						Well-being[b]					
		Total effect		Natural direct effect		Natural indirect effect		Total effect		Natural direct effect		Natural indirect effect	
		RR	95% CI	RR	95% CI	RR	95% CI	RR	95% CI	RR	95% CI	RR	95% CI
		Only income[d]											
Education[c]	High	1.00		1.00		1.00		1.00		1.00		1.00	
	Low	1.12	1.07, 1.18	1.06	1.01, 1.14	1.06	1.01, 1.13	1.03	1.01, 1.06	1.00	0.98, 1.08	1.02	1.01, 1.04
		Only management position[e]											
	High	1.00		1.00		1.00		1.00		1.00		1.00	
	Low	1.08	1.03, 1.15	1.04	1.01, 1.17	1.04	1.01, 1.07	1.00	0.96, 1.04	1.00	0.98, 1.08	0.99	0.95, 1.04
		Only occupational hierarchy position[f]											
	High	1.00		1.00		1.00		1.00		1.00		1.00	
	Low	1.14	1.11, 1.20	1.06	1.02, 1.14	1.08	1.03, 1.13	1.00	0.96, 1.03	1.10	0.98, 1.08	0.98	0.98, 1.03
		Only subjective social status[g]											
	High	1.00		1.00		1.00		1.00		1.00		1.00	
	Low	1.10	1.05, 1.15	1.06	1.01, 1.13	1.04	1.01, 1.09	1.02	0.99, 1.08	1.01	0.98, 1.08	1.01	0.98, 1.01

[a]EQ-5D: Health was assessed by the EQ-5D generic measure of health-related quality of life. Healthy for EQ-5D included all respondents ticking level one for all five dimensions

[b]SWLS: Well-being was measured by the first three items from the satisfaction with life scale measured on a 7-point scale. High wellbeing included those who reported 6 or 7 for all three items..

[c]Education: (Low= primary and secondary school or similar (i.e. 7–10 years of schooling) or vocational school or high school,) (High= college or university (less than 4 years) or college or university (4 years or more))

All models adjusted for potential confounding variables (age, gender, exposure to passive smoke in childhood, having enough friends, number of friends, marital status, physical activity, mothers' psychological problems, fathers' psychological problems, and childhood socioeconomic status.)

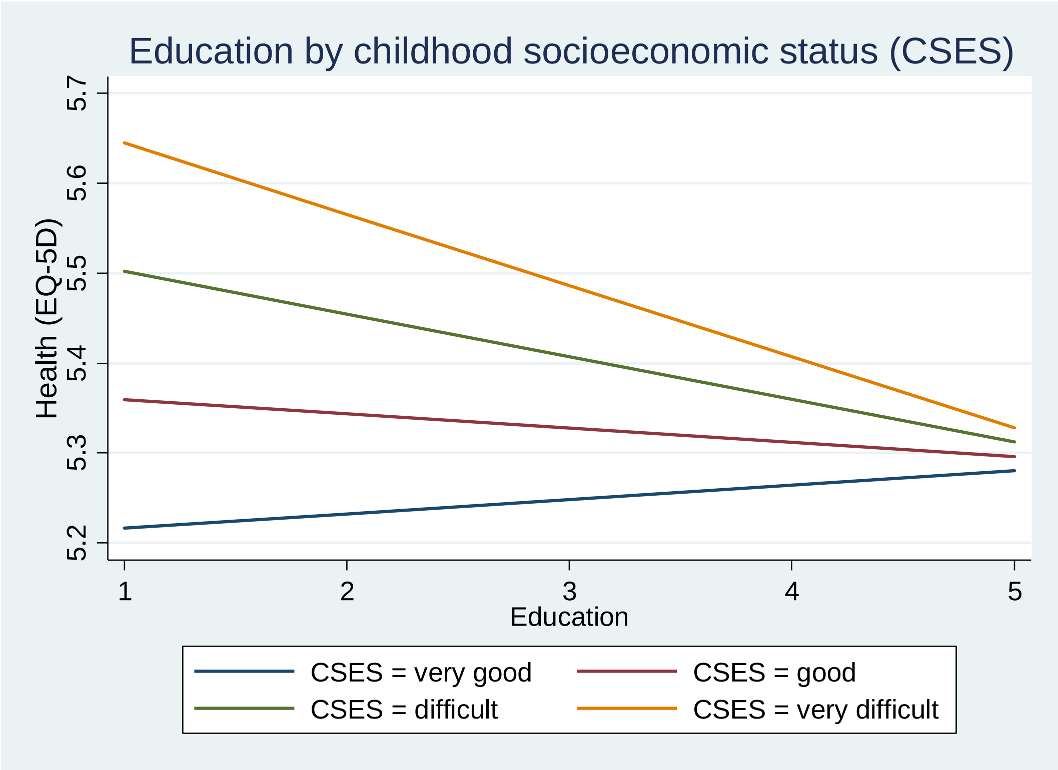
[d]Adjusted for confounding variables + management position, occupational hierarchy position and subjective social status.

[e]Adjusted for confounding variables + income, occupational hierarchy position and subjective social status.

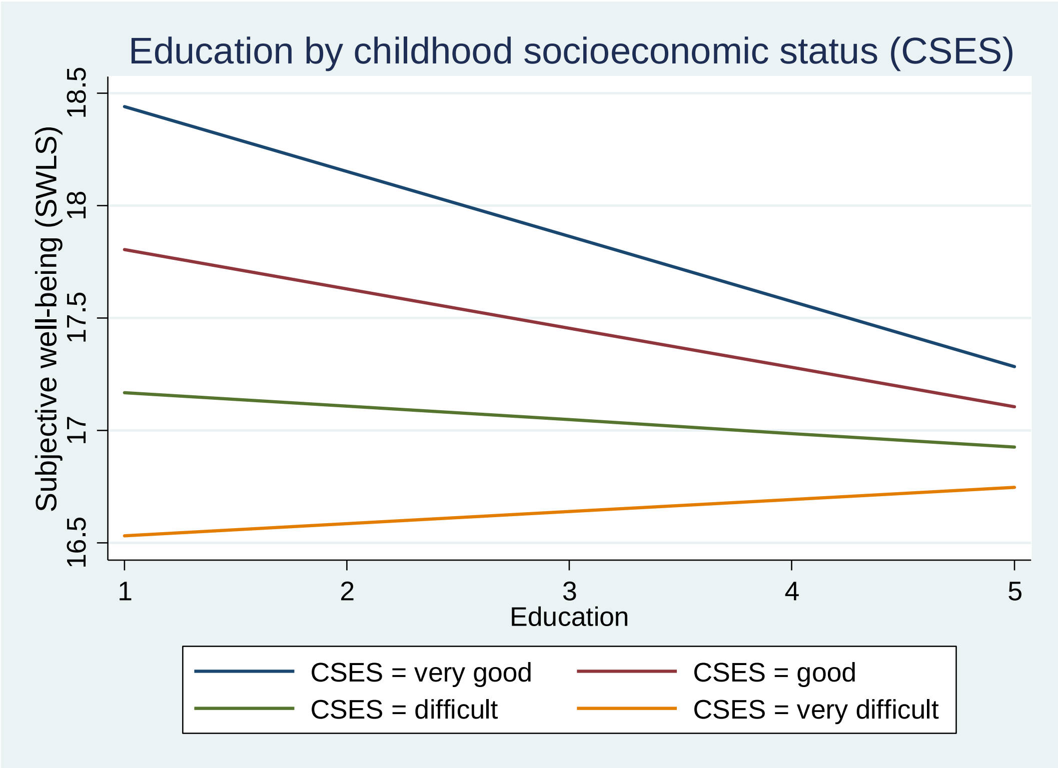
[f]Adjusted for confounding variables + income, management position and subjective social status.

[g]Adjusted for confounding variables + income, management position and occupational hierarchy position.

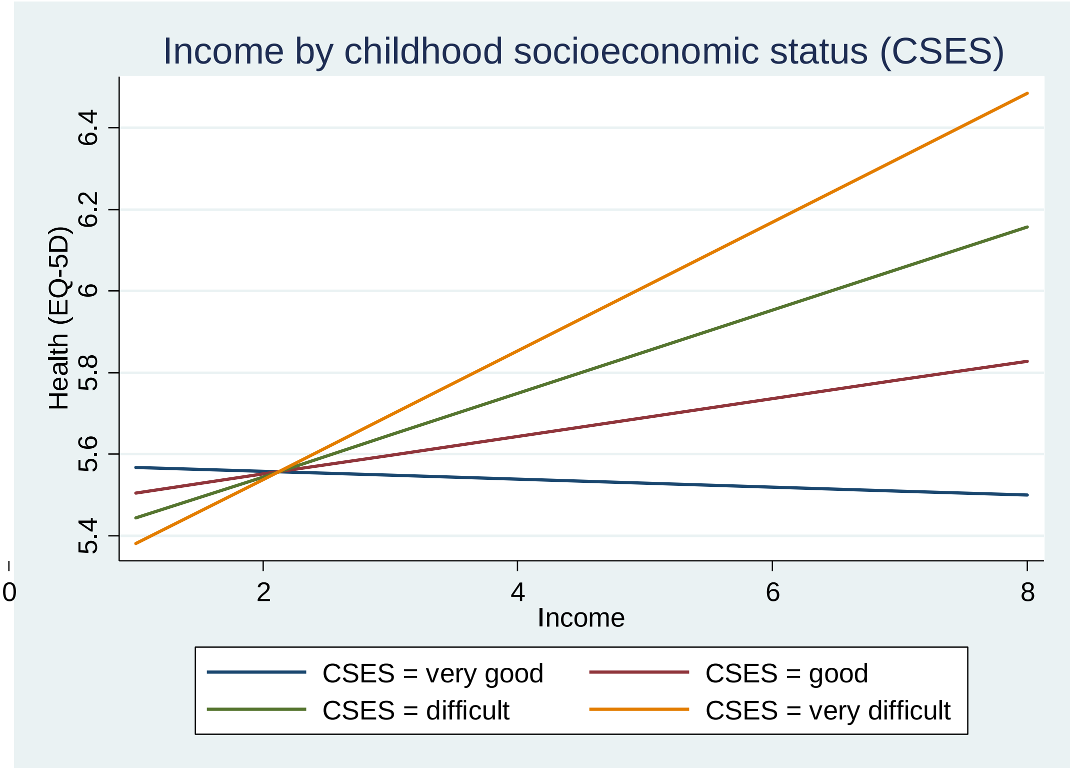
eFigure 1: The effect of education on health by childhood socioeconomic status.



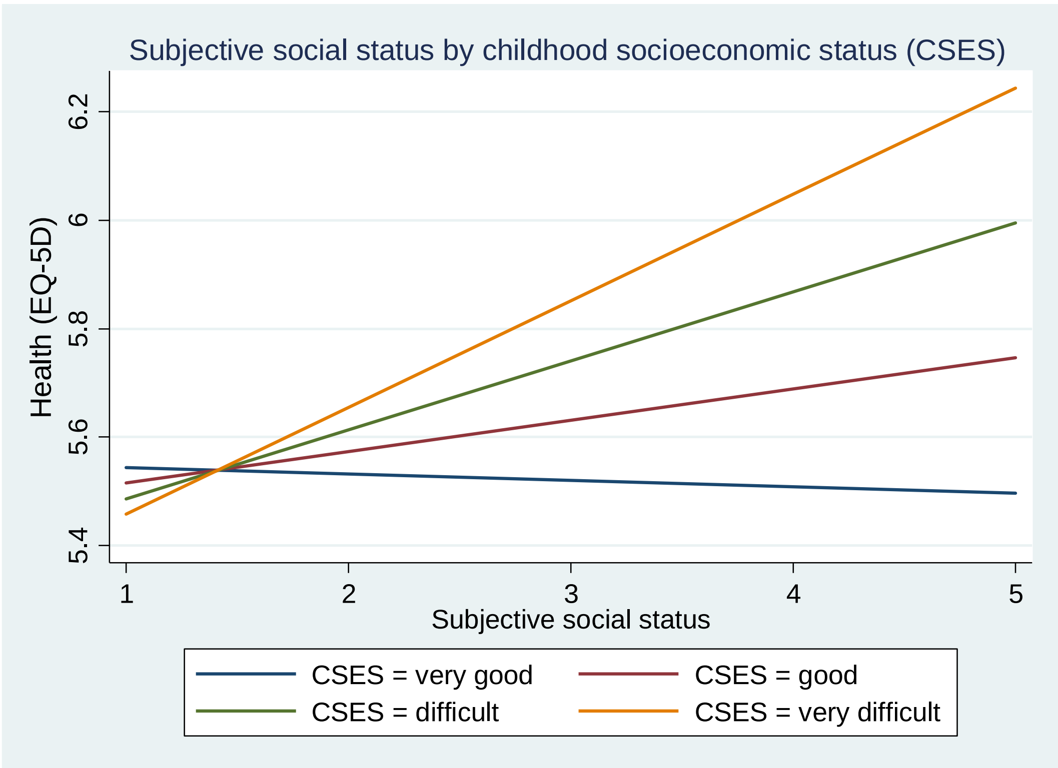
eFigure 2: The effect of education on subjective well-being by childhood socioeconomic status.



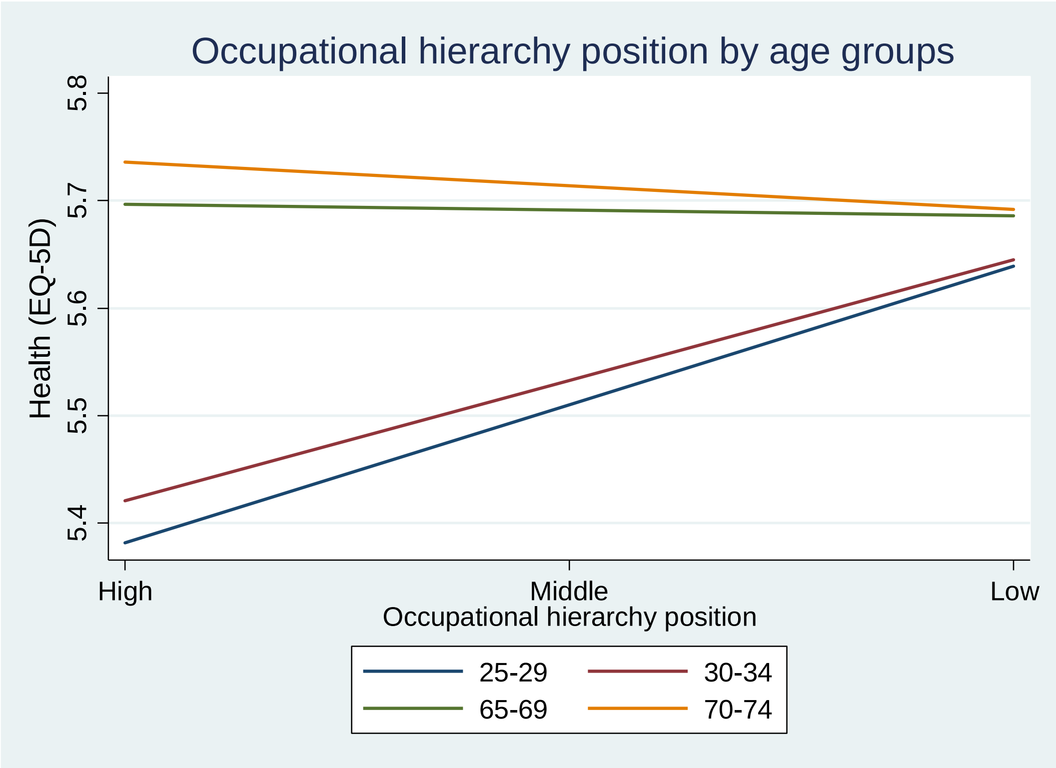
eFigure 3: The effect of income on health by childhood socioeconomic status.



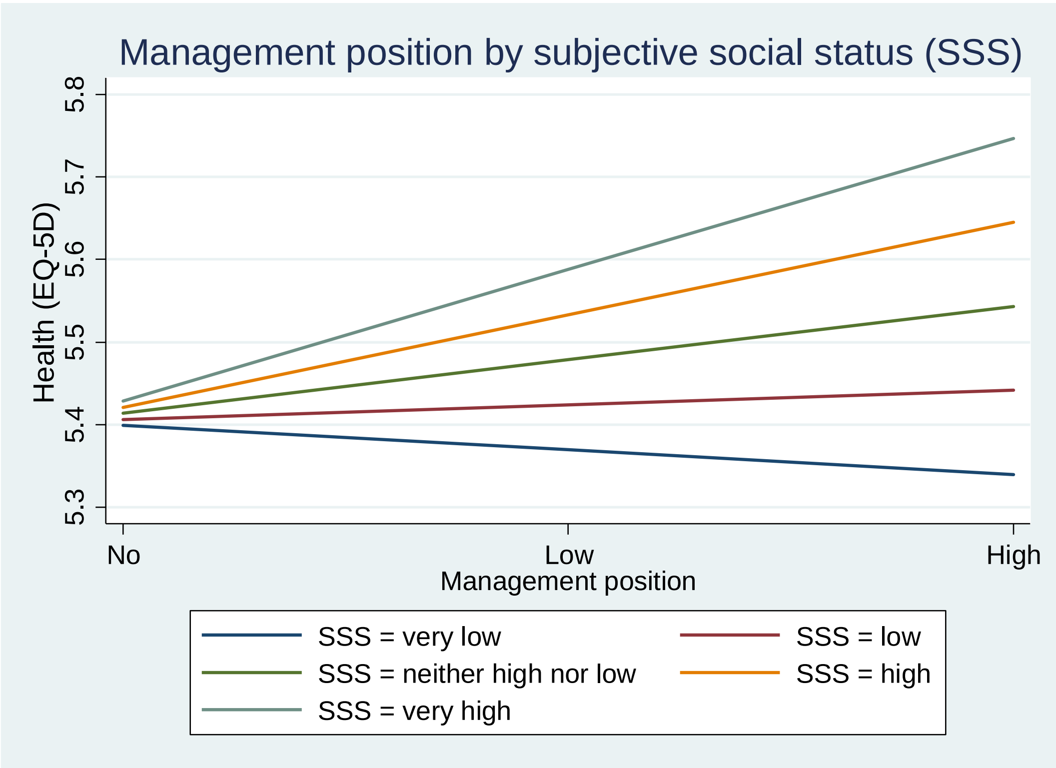
eFigure 4: The effect of subjective social status on health by childhood socioeconomic status.



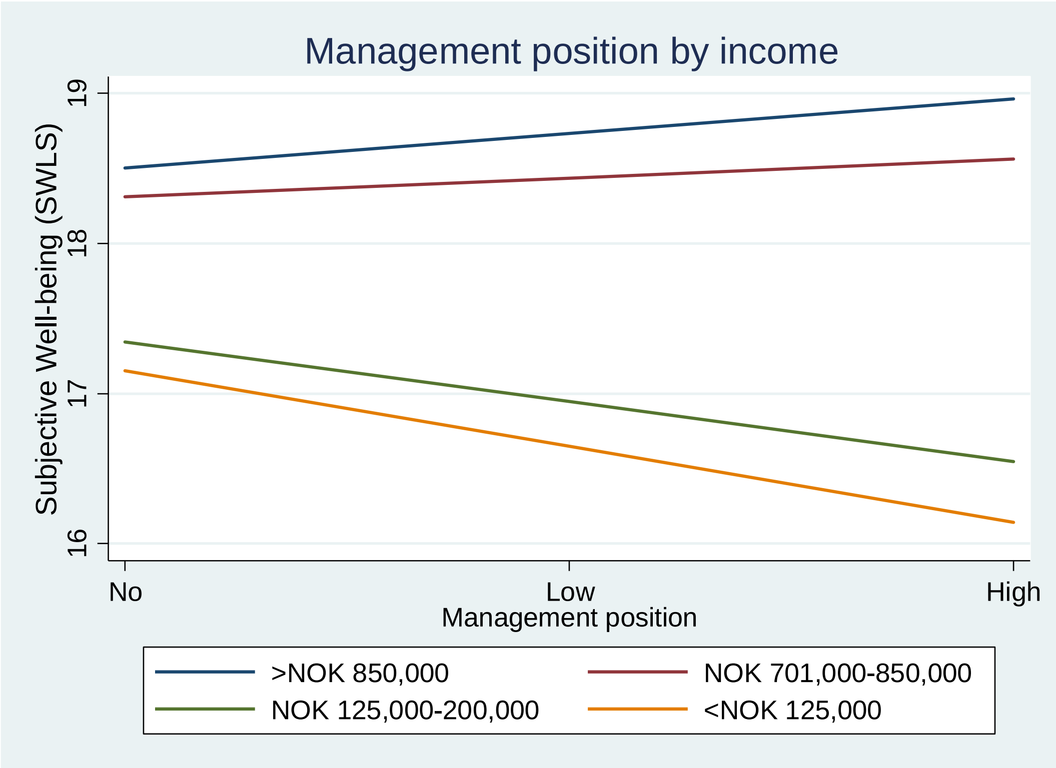
eFigure 5: The effect of occupational hierarchy position on health by age groups.



eFigure 6: The effect of management position on health by subjective social status



eFigure 7: The effect of management position on subjective well-being by income.



eFigure 8: The effect of management position on subjective well-being by occupational hierarchy position.