PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Association of socioeconomic status measured by education, and cardiovascular health: a population-based cross-sectional study
AUTHORS	Jankovic, Slavenka; Stojisavljevic, Dragana; Jankovic, Janko; Eric, Milos; Marinkovic, Jelena

VERSION 1 - REVIEW

REVIEWER	Jing Fang CDC, USA
REVIEW RETURNED	08-Apr-2014

GENERAL COMMENTS	Although the topic had been reported early, it would be an interest report if the purpose is to add information for the Republic of Srpska. However, the study include only education, which affected by other characteristics. I would suggest the authors can include other demographic characteristics (marital status, employment status and type of settlement) and have appropriate adjustment. This population-based survey study at Republic of Srpska showed that cardiovascular health was related to level of education, similar as earlier studies reported in the US. In generally, there is need to improve in English grammar. I have some specific comments for the methods/results of the report: Abstract: Participants – the authors defined CVD as coronary heart disease or stroke. However, CVD should be defined as total cardiovascular disease. If only CHD and stroke, I would suggest using atherosclerotic cardiovascular disease. RS – please spell out when first use Methods – while there was limited information on socioeconomic status, some information was available (marital status, employment status and type of settlement). It would be interest if the authors can explore the cardiovascular health status with these characteristics. Of course, if there were more than one characteristic, the authors should switch the dependent/independent variables. Therefore, the dependent variable would be cardiovascular health status and independent variables would be different sociodemographic characteristics. Methods – it is not clear how the healthy diet score was measured. Please be more specific about it. Also, it is unknown if sodium intake is included in the measurement. I would suggest the authors to provide an appendix of healthy diet score category. Methods – It's still not clear why authors had smoking included in both behavior and health conditions. Results – the final sample of population 4015 or 4165? Results – were any of the reported percentage were adjusted by age

	or sex? if education differed by age and sex, the measured CV health by education should also be adjusted by these variables.

REVIEWER	Linda Ernstsen, PhD Sør-Trøndelag University College, Norway
REVIEW RETURNED	13-Apr-2014

GENERAL COMMENTS	I am familiar with logistic regression, but not with the concept of treating education level as the dependent variable and the CVD health factors as independent variables. I have not seen this strategy used in other studies on social gradients in health either. This is the main reason why I suggest that the paper should be seen by a statistician. This cross-sectional study from the Republic of Srpska investigates the prevalence of cardiovascular (CVD) health metrics (as defined by American Heart Association), and socioeconomic status measured by education. The results show that cardiovascular health is positively related to education. The topic of the manuscript is interesting because it deals with a major public health concern. Still, there are necessary major revisions to be dealt with and they are commented below. Abstract: According to AHA the seven CVD health metrics can be presented in the following ways 1. Ideal CVD health status defined as all seven ideal metrics present (versus immediate or poor CVD health status) 2. Ideal health factor index (all 4 metrics present), and 3. Ideal health behaviors index (all 4 metrics present). The concepts "the ideal health factor index" and "the ideal health behavior index "(with the associated metrics in brackets) should be introduced and used in the abstract, and the three different concepts should be consistently used in the text. The reporting of mean number in odds ratio is unusual (the last sentence in page 2). To be able to initiate an appropriate intervention strategy, politicians need more information about the social distribution of behaviors versus the social distribution of health factors (clinically measured). The conclusions may not be justified by the results reported in the abstract. Introduction: In the last sentence of the introduction, the authors claim that the aim of the study was to investigate SES gradients in the prevalence of CVD metrics. As mentioned for the abstract, the aim (s) could be clearer related to the analyses performed (see table 3 and comments related to statistical analyses). Usually, to detect a social gradient, more than two SES groups should be compared. Methods: The major concern for this study is the study population, and for several reasons: 1. If the study population is representative of the population of R Srpska, it is unlikely that all of the participants were free from CVD at inclusion (as stated in the abstract and in the first sentence in the Result-section). Thus, it seems that National Health Survey constitutes one population and that the sample in the present study is drawn from this. The sampling plan for the National Health Survey should be explained, further, the selection of the study sample (and number excluded) should be described in more details. How many persons living in collective household were excluded? How was "free from CVD" defined, and how many participants with CVD were excluded? The rationale behind excluding those with CVD is 2

unclear, as it seems that that the national goals set by AHA (reference 4 in the reference list) include the whole population.
2. It is highly speculative to include adults at the age of 18 when measuring socioeconomic differences in health for two reasons. Firstly, the cut-off values of each health metric set by AHA are related to persons over 20 years of age. Secondly, education is not established in early childhood as stated in the last paragraph at page 13. Participants of the age of 18 have not been able to finish college or a university degree yet, even if they at the present time have decided to achieve higher education. Further, it is quite common that young adults at this age still live with their parents while being unemployed and/or students. If multiple respondents are living within the same household the regression independence assumption might be violated. The authors should perform sensitivity analyses by excluding those under 25 years of age.
3. Even if there should be too few women to perform stratified regression analyses by gender, baseline data should preferably be presented separately for women and men. It would be useful to see more details about the distribution of age of the study population- what was the mean, median and max age in men and women in the different categories of education? This is of special interest when it comes to interpretations of results/the discussion and, for any choice of public health interventions.
4. A large proportion of the study population was probably retired or disabled (categorised as inactive) when they were interviewed and some of the oldest participants may not had the opportunity to achieve a higher education level when they were young. Therefore, it is plausible that the influence of education (and other social characteristics) on health is systematically different in the study subjects depending on their age (and gender).
I strongly recommend that the authors perform and report results from subgroup analysis stratified by age (or at least for two broad groups, of those in working age and those in the retirement population). It is also plausible that the importance of education in women in particular, could differ for different birth cohorts. These differences cannot be detected if age is “adjusted for” in analyses and all subjects with age ranging from 18 years (or 25 years) years to around 80 years of age.
The study variables and the cardiovascular health status variables, page 6-9, are well described.
Statistical analysis:
Differences in characteristics between participants with different education levels (table 1) are usually analysed by ANOVA for continuous variables and chi-square for categorical variables (or non-parametric tests based on the distribution). 3

In the last sentence at page 9, it is stated that the dependent variable was educational level. From table 3 is also seems as if the CVD health metrics is treated as independent variables and education, as stated, as the dependent variable. Is this correct? If so, the rationale behind this choice must be explained. Especially as most of the studies on the social distribution of CVD risk factors, including those in the reference list, treat education as the independent variable. Alternatively, as ideal cardiovascular health defined by AHA is hardly present (in any populations), poor cardiovascular health could serve as the dependent variable and low educational level could serve as the reference group that the two other educational groups (used as dummy variables in the analysis) are compared to. This would give the reduced odds for not having poor health for those with medium and high education. Or, five or

	more ideal components may serve as the outcome in logistic regression model due to the small number of participants having all 7 ideal components. It seems as age is used as a continuous variable assuming an exactly linear relationship with the outcome in the adjusted analyses. That might not be the case. The most common way of treating age in social epidemiological studies is dividing the ages into categories – often 5- or 10-year age groups. It would also be interesting if the authors could analyse the ideal health factor index (all four or three factors present) and the ideal health behaviour index (all four or three factors present) in two separate analyses with education as the independent variable. Linear regression may also be used if health metrics (as a quantitative variable with a score 0-7) is set as the dependent variable. Results: Table 1 and 2: Why were the comparisons of CVD metrics between medium versus high education not adjusted for sex (as for the other comparisons between educational levels)? In table 3, no. of ideal CV health metrics, the second decimal (hundredth) is missing for those with high education. Discussion: Due to the age range in the present study, the conceptual meaning of education as a proxy for socioeconomic position is violated. This also hampers any comparison of the results with studies using another age range of adult populations. Further, being at the lowest education level is strongly connected to age and gender because of transformation of social structures (number of women in the workforce, technology, the educational system itself e.g.). In light of this, any study of the social distribution of cardiovascular health should preferably stratify analyses by gender. Thus, the authors have made some important considerations regarding the transition from “diseases of affluence” to the diseases of the poor”. Nevertheless, some reflections could also be applied to the transition from “diseases of men” to “diseases including women”.

REVIEWER	Le Kang Virginia Commonwealth University, USA
REVIEW RETURNED	17-Apr-2014

GENERAL COMMENTS	The authors investigated association between socioeconomic status and cardiovascular health in the adult population of the Republic of Srpska based on a cross-sectional study. The article is generally easy to follow. Still, I have the following critics. As authors themselves pointed out, the consideration of education only for socioeconomic status limits the scope of this study. Can you justify the reason why you don’t consider other factors, such as professions? Unless you could show that the mean number of ideal cardiovascular metrics follow normal distribution, the reporting of mean, standard errors, 95% CI or hypothesis testing does not fully justify the conclusion. For instance, why not median number? Have you ever considered the ordinal logistic regression as you have low, medium and high for education levels? ***Minor:*** On page 10, line 41-42, the authors stated “most participants had

	ideal levels of cardiovascular health metrics….” On page 11, line 10-11, the authors stated “virtually no one had ideal cardiovascular health…” It is confusing and authors need to explain more details, not just pointing to different tables.

VERSION 1 – AUTHOR RESPONSE

RESPONSES TO THE REVIEWER 1 (JING FANG)

1. Thank you to give me the opportunity to review this manuscript. Although the topic had been reported early, it would be an interest report if the purpose is to add information for the Republic of Srpska. However, the study include only education, which affected by other characteristics. I would suggest the authors can include other demographic characteristics (marital status, employment status and type of settlement) and have appropriate adjustment.

RESPONSE: We accepted your suggestions. Marital status, employment status and type of settlement have been entered as covariates in the multivariate logistic regression analysis where appropriate (based on the results in Table 1). Please see the new Table 3. In the Method section, subtitle Statistical analysis we added marital status, employment status and type of settlement as adjusters.

2. This population-based survey study at Republic of Srpska showed that cardiovascular health was related to level of education, similar as earlier studies reported in the US.

RESPONSE: That is true.

3. In generally, there is need to improve in English grammar

RESPONSE: The language has been improved.

4. I have some specific comments for the methods/results of the report:
4a. Abstract: Participants – the authors defined CVD as coronary heart disease or stroke. However, CVD should be defined as total cardiovascular disease. If only CHD and stroke, I would suggest using atherosclerotic cardiovascular disease.

RESPONSE: According to the comments of the reviewer 2 we deleted the wording: "to be free of CVD (coronary heart disease or stroke)", from the Abstract.

4b. RS – please spell out when first use

RESPONSE: We did.

4c. Methods – while there was limited information on socioeconomic status, some information was available (marital status, employment status and type of settlement). It would be interest if the authors can explore the cardiovascular health status with these characteristics. Of course, if there were more than one characteristic, the authors should switch the dependent/independent variables. Therefore, the dependent variable would be cardiovascular health status and independent variables would be different socio-demographic characteristics.

RESPONSE: Our hypothesis was that different indices of cardiovascular health in this population survey are classificators / discriminators / explorators of educational levels (i.e. socioeconomic

status). It means that the logistic regression analysis was applied as classification (and not prediction) tool [Afifi A, May S, Clark V. Computer-Aided Multivariate Analysis, Fourth Edition. Chapman & Hall / CRC Texts in Statistical Science, 2004]. It has been already mentioned in the Method section. In light of this fact adding other socioeconomic variables does not imply the dependent/independent variables switch. We added suggested variables as new “independent” variables in the multivariate logistic regression analysis (new Table 3). Consequently, all necessary changes were made in the Result section.

4d. Methods – it is not clear how the healthy diet score was measured. Please be more specific about it. Also, it is unknown if sodium intake is included in the measurement. I would suggest the authors to provide an appendix of healthy diet score category.

RESPONSE: In the Method section we have now described in detail how the healthy diet score was measured. Sodium intake was included in the measurement. The categories of healthy diet score have now been provided in the Appendix 1.

4e. Methods – It’s still not clear why authors had smoking included in both behavior and health conditions.

RESPONSE: This is according to AHA’s definition [Lloyd-Jones et al. Defining and Setting National Goals for Cardiovascular Health Promotion and Disease Reduction The American Heart Association’s Strategic Impact Goal Through 2020 and Beyond. Circulation. 2010; 121: 586-613]: "Given the importance of abstinence from smoking and smoking cessation to health promotion, smoking appears in both lists of health behaviours and health factors" (2).

4f. Results – the final sample of population 4015 or 4165?

RESPONSE: Both. We started with 4165 for description of population and specific dimensions of cardiovascular health (CVH). To analyse indexes or categories of CVH we have used 4015 participants who had data on all specific dimensions of CVH. This is already mentioned in the Method section, subtitle Statistical analysis: "The analyses were carried out on 4165 participants except when indices of ideal CVH were in question. These analyses were done on 4015 participants. We excluded 150 subjects who lacked data on any of the CVH health metrics".

4g. Results – were any of the reported percentage were adjusted by age or sex? If education differed by age and sex, the measured CV health by education should also be adjusted by these variables.

RESPONSE: All reported percentages (the prevalence data) were adjusted by age and sex. Now we have marked with footnotes that the prevalences in Table 2 and 3 are adjusted by age and sex.

Again, we appreciate all your insightful comments.

RESPONSES TO THE REVIEWER 2 (LINDA ERNSTSEN)

I am familiar with logistic regression, but not with the concept of treating education level as the dependent variable and the CVD health factors as independent variables. I have not seen this strategy used in other studies on social gradients in health either. This is the main reason why I suggest that the paper should be seen by a statistician.
This cross-sectional study from the Republic of Srpska investigates the prevalence of cardiovascular (CVD) health metrics (as defined by American Heart Association), and socioeconomic status measured by education. The results show that cardiovascular health is positively related to education. The topic of the manuscript is interesting because it deals with a major public health concern. Still,

there are necessary major revisions to be dealt with and they are commented below.
Abstract: According to AHA the seven CVD health metrics can be presented in the following ways 1. Ideal CVD health status defined as all seven ideal metrics present (versus immediate or poor CVD health status) 2. Ideal health factor index (all 4 metrics present), and 3. Ideal health behaviors index (all 4 metrics present). The concepts “the ideal health factor index” and “the ideal health behavior index” (with the associated metrics in brackets) should be introduced and used in the abstract, and the three different concepts should be consistently used in the text.

RESPONSE: The concepts “the ideal health factor index” and “the ideal health behavior index” (with the associated metrics in brackets) now have been introduced in the abstract.

The reporting of mean number in odds ratio is unusual (the last sentence in page 2). To be able to initiate an appropriate intervention strategy, politicians need more information about the social distribution of behaviors versus the social distribution of health factors (clinically measured). The conclusions may not be justified by the results reported in the abstract.

RESPONSE: The reporting of mean number in odds ratio is unusual, but it is possible. We have changed the results and conclusion in the abstract. The conclusions now may be justified by the results reported in the abstract.

Introduction: In the last sentence of the introduction, the authors claim that the aim of the study was to investigate SES gradients in the prevalence of CVD metrics. As mentioned for the abstract, the aim (s) could be clearer related to the analyses performed (see table 3 and comments related to statistical analyses). Usually, to detect a social gradient, more than two SES groups should be compared.

RESPONSE: We have changed the aim of the study, and we hope that it is now more clearly related to the analyses performed.

Methods: The major concern for this study is the study population, and for several reasons:
1. If the study population is representative of the population of R Srpska, it is unlikely that all of the participants were free from CVD at inclusion (as stated in the abstract and in the first sentence in the Result-section). Thus, it seems that National Health Survey constitutes one population and that the sample in the present study is drawn from this. The sampling plan for the National Health Survey should be explained, further, the selection of the study sample (and number excluded) should be described in more details. How many persons living in collective household were excluded? How was “free from CVD” defined, and how many participants with CVD were excluded? The rationale behind excluding those with CVD is unclear, as it seems that the national goals set by AHA (reference 4 in the reference list) include the whole population.

RESPONSE: The sampling plan for the National Health Survey, the selection of the study sample are now described in more details in the Method section. The fact is that the participants with CVD were not excluded. That’s why we deleted the wording "to be free of CVD (coronary heart disease or stroke)" from the Abstract and from the Method section, subtitle Study design and participants. The analyses were carried out on 4165 participants except when indices of ideal cardiovascular health were in question. These analyses were done on 4015 participants. We excluded 150 subjects who lacked data on any of the cardiovascular health metrics.
According to the official data the estimated number of persons living in collective household (refugees after the war in the former Yugoslavia) in Republic of Srpska in 2010 was about 250,000.

2. It is highly speculative to include adults at the age of 18 when measuring socioeconomic differences in health for two reasons. Firstly, the cut-off values of each health metric set by AHA are related to persons over 20 years of age. Secondly, education is not established in early childhood as

stated in the last paragraph at page 13. Participants of the age of 18 have not been able to finish college or a university degree yet, even if they at the present time have decided to achieve higher education. Further, it is quite common that young adults at this age still live with their parents while being unemployed and/or students. If multiple respondents are living within the same household the regression independence assumption might be violated. The authors should perform sensitivity analyses by excluding those under 25 years of age.

RESPONSE: We agree that is speculative to include adults at the age of 18 when measuring socioeconomic differences in health. We have changed the wording about education on page 13 and now it is stated that education is established relatively early in life, and it is stable over the adult life span.
As you suggested we performed sensitivity analyses by excluding those under 25 years of age (please see Appendix 2). The results presented in Appendix 2 have now been described in the Result section.

3. Even if there should be too few women to perform stratified regression analyses by gender, baseline data should preferably be presented separately for women and men. It would be useful to see more details about the distribution of age of the study population - what was the mean, median and max age in men and women in the different categories of education? This is of special interest when it comes to interpretations of results/the discussion and, for any choice of public health interventions.

RESPONSE: The distribution of age of the study population (mean, median, min and max) by gender in the different categories of education are now presented in the Legend of Table 1. We also performed the stratified regression analysis by gender (see Appendix 2).

4. A large proportion of the study population was probably retired or disabled (categorized as inactive) when they were interviewed and some of the oldest participants may not had the opportunity to achieve a higher education level when they were young. Therefore, it is plausible that the influence of education (and other social characteristics) on health is systematically different in the study subjects depending on their age (and gender).
I strongly recommend that the authors perform and report results from subgroup analysis stratified by age (or at least for two broad groups, of those in working age and those in the retirement population). It is also plausible that the importance of education in women in particular, could differ for different birth cohorts. These differences cannot be detected if age is “adjusted for” in analyses and all subjects with age ranging from 18 years (or 25 years) years to around 80 years of age.

RESPONSE: We performed a subgroup analysis stratified by age (Appendix 2) and gender and reported results in the Result section. In general, the relationship between the number of ideal CVH metrics and CVH categories, and educational levels did not differ significantly in different age subgroups (The only exception was for study participants ≥25 years in low versus high educational group). We agree that the importance of education in women could differ for different birth cohorts. However we think that this overcomes the aims of this paper.

The study variables and the cardiovascular health status variables, page 6-9, are well described.

RESPONSE: Thank you.

Statistical analysis:
Differences in characteristics between participants with different education levels (table 1) are usually analysed by ANOVA for continuous variables and chi-square for categorical variables (or non-parametric tests based on the distribution).

RESPONSE: This is true. But concerning that any regression analysis provides more information over any statistical test – the size or measure of a difference (if any) in the form of partial regression coefficients, we decided to use the latter. However we did the suggested tests and added the results in Table 1.
In the last sentence at page 9, it is stated that the dependent variable was educational level. From table 3 is also seems as if the CVD health metrics is treated as independent variables and education, as stated, as the dependent variable. Is this correct? If so, the rationale behind this choice must be explained. Especially as most of the studies on the social distribution of CVD risk factors, including those in the reference list, treat education as the independent variable. Alternatively, as ideal cardiovascular health defined by AHA is hardly present (in any populations), poor cardiovascular health could serve as the dependent variable and low educational level could serve as the reference group that the two other educational groups (used as dummy variables in the analysis) are compared to. This would give the reduced odds for not having poor health for those with medium and high education. Or, five or more ideal components may serve as the outcome in logistic regression model due to the small number of participants having all 7 ideal components.
RESPONSE: These two paragraphs need somewhat longer explanation. Our hypothesis was that different metrics of cardiovascular health are classificators / discriminators / explorators of educational levels (i.e. socioeconomic status) in this population survey. (Of course, it could be true not for all levels but only for some of them, and this was the case in our results, i.e. cardiovascular health indices were disriminators between high and low+medium educational level participants.) This means that logistic regression analysis was applied as classification (and not prediction) tool – more exactly as supervised classification tool (Afifi A, May S, Clark V. Computer-Aided Multivariate Analysis, Fourth Edition. Chapman & Hall / CRC Texts in Statistical Science, 2004). In the light of this usage, even the numerical part is the same, dependent and independent does not have the same meaning as in prediction problems. Dependent is some kind of structured groups and independent variables are possible classificators / discriminators / explorators of those groups. Truly, this is not frequently the case in this kind of studies but in the other fields it is. We hope to see it more in social and epidemiological studies.
As reviewer 1 suggested, besides sex and age, marital status, employment status and type of settlement have been entered as covariates in the multivariate logistic regression analysis where appropriate (based on the results in Table 1). Please see the new Table 3. The final results have been changed a little bit.

It seems as age is used as a continuous variable assuming an exactly linear relationship with the outcome in the adjusted analyses. That might not be the case. The most common way of treating age in social epidemiological studies is dividing the ages into categories – often 5- or10-year age groups.

RESPONSE: We divided age in 10-year age groups but for the sake of space we used mean and SD as its descriptors.

It would also be interesting if the authors could analyse the ideal health factor index (all four or three factors present) and the ideal health behaviour index (all four or three factors present) in two separate analyses with education as the independent variable. Linear regression may also be used if health metrics (as a quantitative variable with a score 0-7) is set as the dependent variable.

RESPONSE: Both of these suggested analyses were performed in another paper (Int J Public Health, 2014 in press).

Results: Table 1 and 2: Why were the comparisons of CVD metrics between medium versus high education not adjusted for sex (as for the other comparisons between educational levels)?

RESPONSE: Because the sex was not the discriminator between participants with medium and high educational level (see Table 1).

In table 3, no. of ideal CV health metrics, the second decimal (hundredth) is missing for those with high education.

RESPONSE: We added the second decimal.

Discussion: Due to the age range in the present study, the conceptual meaning of education as a proxy for socioeconomic position is violated. This also hampers any comparison of the results with studies using another age range of adult populations. Further, being at the lowest education level is strongly connected to age and gender because of transformation of social structures (number of women in the workforce, technology, the educational system itself e.g.). In light of this, any study of the social distribution of cardiovascular health should preferably stratify analyses by gender. Thus, the authors have made some important considerations regarding the transition from “diseases of affluence” to the diseases of the poor”. Nevertheless, some reflections could also be applied to the transition from “diseases of men” to “diseases including women”.

RESPONSE: According to the results of sensitivity analyses stratified by age (Appendix 2) we think that the conceptual meaning of education as a proxy for socioeconomic position in our study is not significantly violated. In the light of new findings from the analysis stratified by gender (Appendix 2) we added one paragraph related to association between sex and education in the Discussion section. Concerning comparison of our results with studies using another age range of adult populations this is not case for all studies we used for comparison. For example we used the same range of adult population (≥18 years) like Graciani et al. (Circ Cardiovasc Qual Outcomes 2013;6:90–8 –our reference 13) and Fang et al. (J Am Heart Assoc 2012; 1:e005371. doi: 10.1161/JAHA.112.005371 – our reference 12) in their studies. We have changed the conclusions a little bit.

Again, we appreciate all your insightful comments. We worked hard to be responsive to them. Thank you for taking the time and energy to help us improve the paper.

RESPONSES TO THE REVIEWER 3 (LE KANG)

The authors investigated association between socioeconomic status and cardiovascular health in the adult population of the Republic of Srpska based on a cross-sectional study. The article is generally easy to follow. Still, I have the following critics.
1. As authors themselves pointed out, the consideration of education only for socioeconomic status limits the scope of this study. Can you justify the reason why you don’t consider other factors, such as professions?

RESPONSE: Education was selected as our indicator of socioeconomic position, given that it is attained relatively early in life and it is stable over the adult life span. In the INTERHEART case-control study, education as a proxy for socioeconomic position was found to be most consistently associated with increased risk for acute myocardial infarction globally [Rosengren A, Subramanian SV, Islam S, Chow CK, Avezum A, Kazmi K, Sliwa K, Zubaid M, Rangarajan S, Yusuf S: Education and risk for acute myocardial infarction in 52 high, middle and low-income countries: INTERHEART case-control study. Heart 2009, 95(24):2014–2022.1]. Further, most studies confirm a significant link between education and health behaviour, and the association between education and IHD mortality is stronger than occupational position or income-based measures [Marmot M, Bartely M: Social class and coronary heart disease. In Stress and the heart: Psychosocial pathways to coronary heart disease. Edited by Stansfeld S, Marmot M. London: BMJ Books; 2002:5–19.2].
Besides, the profession in our study was recorded in less than a half of all participants. Employment

status was treated as covariate even its association with education is considerably high (see the Table below) as expected.

Symmetric Measures

	Value	Asymp. Std. Error	Approx. Tb	Approx. Sig.
Kendall's tau-b	-.208	.013	-15.570	.000
Kendall's tau-c	-.190	.012	-15.570	.000
Gamma	-.319	.020	-15.570	.000
N of Valid Cases	4162			

a. Not assuming the null hypothesis.
b. Using the asymptotic standard error assuming the null hypothesis.

2. Unless you could show that the mean number of ideal cardiovascular metrics follow normal distribution, the reporting of mean, standard errors, 95% CI or hypothesis testing does not fully justify the conclusion. For instance, why not median number?

RESPONSE: Even if distribution is not normal (according to Kolmogorov-Smirnov test it is not) they are at least symetric and based on large samples and it should not compromise the analysis. We calculated the median values and IQRs and put them in Table 3. Along with it we added, besides age and sex, other socio-demographic covariates where appropriate - as suggested by reviewer 1. The results, as expected, have now changed a bit.

3. Have you ever considered the ordinal logistic regression as you have low, medium and high for education levels?

RESPONSE: Yes we did. The reasons why we have chosen binary logistic regressions for presentation of our results were the following:
Our hypothesis was that different indices of cardiovascular health are clasificators / discriminators / explorators of educational levels (i.e. socioeconomic status) in this population survey. Of course, it could be true not for all levels but only for some of them, and this was the case in our results. I.e. cardiovascular health indices were disriminators between high and low+medium educational level participants.
The most popular and well understood method for ordinal data is the proportional odds model. Our odds were not proportional and the POM is inappropriate. A simple and valid approach to analyze such data is to dichotomize the ordinal response variable by means of several cut-offs and use separate binary logistic regression models for each dichotomized response [Bender R and Grouven U. Using Binary Logistic Regression Models for Ordinal Data with Non-proportional Odds. J Clin Epidemiol Vol. 51, No. 10, pp. 809–816, 1981]. We didn't try to use other ordinal regression models. At the end, the binary logistic regression is more familiar to the audience because it leads to the most straightforward interpretation.
But thank you, we should, attempt to show in other paper the results and concerns associated with ordinal logistic regression analysis.

4. Minor:
On page 10, line 41-42, the authors stated “most participants had ideal levels of cardiovascular health metrics….”
On page 11, line 10-11, the authors stated “virtually no one had ideal cardiovascular health…” It is confusing and authors need to explain more details, not just pointing to different tables.

RESPONSE: These are two different things. Although most participants had ideal levels of

cardiovascular health metric, except for diet and blood pressure, virtually no one had ideal cardiovascular health (i.e. no one had all 7 cardiovascular health metrics at ideal level).

Again, we appreciate all your insightful comments.

VERSION 2 – REVIEW

REVIEWER	Jing Fang CDC, USA
REVIEW RETURNED	13-Jun-2014

GENERAL COMMENTS	The authors addressed most of my comments. However, I'm still not sure if it is appropriate to use education as dependent variable in logistic regression model. By this, the authors couldn't identify other important socioeconomic characteristics related to CV health. I would recommend a statistician reviewing this manuscript. If the statistician review conclude that the education as dependent variable and CV metric as independent variable is appropriate, I would recommend to publish the paper.

REVIEWER	Linda Ernstsen Sør-Trøndelag Univrsity College, Norway
REVIEW RETURNED	18-Jun-2014

GENERAL COMMENTS	The authorhs have pointed out in the abstract (and justified it in the respons to the reviewers) that the objective of choosing education as the dependent variable was to assess wether indices of CVH were discriminators of socioeconomic status. Not many studies have assessed the association between CVH metrics and socioeconomic position in this way, so it might be possible that readers literally mix up the CHV metrics as the dependent variables as they are presented under the heading "outcome measures" in the abstract. Or perhaps this is of minor concern. The authors have met the comments from all of the reviewers, and the paper is markedly improved after revision. I only have two minor suggestions. The first is on the description of the dietary assessment on page 7, line 5 (in the revised version where any changes to the document are accepted). It is stated that the 25-item questionnaire of dietary intake is valdated. If so, there should be given a reference making it possible for any reader to evaluate the validation of this measurement. If there is no published paper on the validation process the word "validated" should not be used (and this sholud rather be mentioned as a limitation of the study). The other comment is on the second last sentence on page 10, line 3. Could the sentence "Most participants had ideal levels of CVH metrics, except for diet and blood pressure" be more precise? When looking at the prevalence of each ideal health metric (Table 2) it is only smoking (54%) and fasting blood glucose (76%) who reach 50% or more (two of the seven metrics), thus using the phrase "most participants" in this case is misleadning. If the prevalence had been 48-49% for the other CVH metrics (except for diet and blood-pressure) the statement would have been more "correct"

REVIEWER	LE KANG Virginia Commonwealth University, USA
REVIEW RETURNED	18-Jun-2014

- The reviewer completed the checklist but made no further comments.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 2
Name Linda Ernstsen
Institution and Country Sør-Trøndelag Univrsity College, Norway
Please state any competing interests or state 'None declared': None declared

The authors have pointed out in the abstract (and justified it in the response to the reviewers) that the objective of choosing education as the dependent variable was to assess whether indices of CVH were discriminators of socioeconomic status. Not many studies have assessed the association between CVH metrics and socioeconomic position in this way, so it might be possible that readers literally mix up the CHV metrics as the dependent variables as they are presented under the heading "outcome measures" in the abstract. Or perhaps this is of minor concern.

RESPONSE: We replaced the heading "Outcome measures" into "Study variables". The next text has been added under this heading: Participant's education was a proxy for SES. Potential discriminators of SES were indices of CVH presented according to AHA as ...

The authors have met the comments from all of the reviewers, and the paper is markedly improved after revision. I only have two minor suggestions. The first is on the description of the dietary assessment on page 7, line 5 (in the revised version where any changes to the document are accepted). It is stated that the 25-item questionnaire of dietary intake is validated. If so, there should be given a reference making it possible for any reader to evaluate the validation of this measurement. If there is no published paper on the validation process the word "validated" should not be used (and this should rather be mentioned as a limitation of the study).

RESPONSE: We added two references (now 9 and 10).

The other comment is on the second last sentence on page 10, line 3. Could the sentence "Most participants had ideal levels of CVH metrics, except for diet and blood pressure" be more precise? When looking at the prevalence of each ideal health metric (Table 2) it is only smoking (54%) and fasting blood glucose (76%) who reach 50% or more (two of the seven metrics), thus using the phrase "most participants" in this case is misleading. If the prevalence had been 48-49% for the other CVH metrics (except for diet and blood-pressure) the statement would have been more "correct".

RESPONSE: The sentence "Most participants had ideal levels of CVH metrics, except for diet and blood pressure" has been changed and now it states: "Relative majority of participants (41.2%–76.0%) had ideal levels of CVH metrics, except for diet (4.4%) and blood pressure (14.7%)."