

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.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Patterns of health-related behaviours among adolescents: a cross-sectional study based on the National Survey of School Health Brazil 2012.
AUTHORS	Azeredo, Catarina; Levy, Renata; Peres, Maria Fernanda; Menezes, Paulo; Araya, Ricardo

VERSION 1 - REVIEW

REVIEWER	Helen Sweeting MRC/CSO Social & Public Health Sciences Unit, University of Glasgow, UK
REVIEW RETURNED	15-Mar-2016

GENERAL COMMENTS	<p>Patterns of health-related behaviours among Brazilian adolescents. Bmjopen-2016-011571</p> <p>Comments for the authors.</p> <p>Thank you for asking me to review this paper which presents an analysis of data on a wide range of health-related behaviours obtained from a sample of over 100,000 Brazilian 9th grade students (83% of those attending school). Gender differences in behaviours are described, factor analysis conducted, resulting in three factors, and the associations between these factors and a number of socio-demographic characteristics are presented.</p> <p>The paper is generally well written and none of my comments is major; most are matters of detail or grammar.</p> <p>Abstract P3, line 22 – suggest ‘... associated WITH each one ...’ P3, line 40 – because the term public school does not mean the same in all countries, suggest that here it says ‘public (compared with private) schools’</p> <p>Strengths/limitations Final bullet – does not make sense, as it says ‘a SINGLE, self-reported behaviour’ when this is analysis of multiple behaviours (same issue in the Discussion (p19, line 47)).</p> <p>Introduction P5, line 6 – suggest ‘health-COMPROMISING ...’ (rather than related), since next sentence says these all cause morbidity/mortality, whereas some health-RELATED behaviours (eg exercise) are health promoting. P5, line 19 – suggest ‘however, some ISSUES remain ...’, since the</p>
-------------------------	--

	<p>points listed are not questions. P5, from line 51 – suggest ‘The factors that shape patterns OF behaviours ... frameworks, MAY VARY BETWEEN high ... P6, line 7 – suggest cutting second ‘of’ (‘... study how ...’).</p> <p>Methods I assume description of the sampling, stratification, etc (p6-7) and analysis (p8-9) are correct and appropriate – my stats isn’t sophisticated enough to be confident on this. P7, final line – did the questionnaire really ask about occupation/ I thought these were school pupils. P8, line 23 – refers to Suppl file – should this be Appendix 1? P8 – categorisation of geographic regions as more/less developed – is this a standard grouping for Brazil? Is there a reference? P8-9 – Descriptive analysis and EFA were both conducted according to gender. That’s fine, but needs some justification, otherwise some readers might wonder why the authors weren’t (also) interested in doing this by age, SES, etc. If gender is a focus, then I think a rationale needs to be provided in the Introduction (this would not be difficult). In addition, having noted separate male and female EFA, the remainder of the analyses does not mention analysis by gender. As a reader, my guess was that this was because there were no significant differences, but it would have been good to know this was correct. P9, lines 36-40 – is this detail on variance explained needed? P10 – multiple imputation is mentioned. Were analyses conducted on the raw and the imputed datasets? Which results are presented?</p> <p>Results P11, first results para – I think a table providing sample characteristics would have been useful. This could also show numbers with missing data for each of these socio-demographic variables. P11, lines 35-37 – males and aggressive behaviour – this is four variables, three dichotomous and one multi-category, so ‘6 percentage points higher’ is unclear. P11, lines 44-46 - I do not think ‘the lowest frequencies of protective behaviours represented a small proportion of students’ is correct. One or no days per week for raw veg was reported by 45%, cooked veg 57%, fruit 38% and physical activity 45%. P12-13 – description of clustering. I think these are factors (first, second, etc), not patterns (they represent an overall pattern). And I wondered slightly about the terms chosen. The second factor is ‘health compromising’, but the first (‘problem behaviours’) is also health compromising. Perhaps the second and third could be ‘health compromising diet and sedentary behaviours’ and ‘health promoting diet and physical activity’ – that’s clearer, even if rather more wordy. P15, Figure 1 – uses ‘HealthY PROMOTING’ whereas the rest of the paper (I think) uses ‘health-enhancing’ – I suggest consistent use of terms. It also includes ‘unsafe sex’, whereas Appendix 1 has this as a heading, not an individual item. P15, line 12 - ‘were weakly correlated’ – these (weak) correlations are not shown on Fig 1, so readers do not know how weak they are. P16, Table 2 – something has gone a bit wrong with the row alignment on the problem behaviour and maternal education results.</p> <p>Discussion P17, 3rd para, 1st sentence – unclear. P18, lines 10-17 – could this sentence be cut? It does not seem to relate to the rest of the paragraph.</p>
--	---

	<p>P18, line 30 – ‘gender differences in behaviours have been decreasing ...’ – this is a very general statement – have differences in ALL behaviours decreased?</p> <p>P18, line 37 - ‘some similarities and differences were found’ – unclear – similarities and differences between what?</p> <p>P19, line 21 – how is SES conceptualised here? In material terms (not measured in the study) or in terms of educational capital (measured via maternal education)? I think this would be worth an extra sentence.</p> <p>P19, line 47 – single behaviour issue, noted earlier.</p> <p>P20, 2nd para – in thinking about interventions, it might also be worth noting prevention activities, which would occur at younger ages, rather than reduction or cessation interventions with older adolescents.</p> <p>Supplementary files 2 and 3 – I would have found an extra column, with significance of the gender difference, useful, because these are quite large tables and it is difficult to assess quickly where the main differences are. A further column could present numbers with missing data for each behavioural variable.</p>
--	---

REVIEWER	Martin Lindström Department of Clinical Sciences in Malmö Lund University Sweden
REVIEW RETURNED	27-Apr-2016

GENERAL COMMENTS	<p>Manuscript: "Patterns of health-related behaviours among Brazilian adolescents"</p> <p>The cross-sectional study from Brazil investigates the clustering of multiple health-related behaviours among adolescents and socio-demographic characteristics associated with these clusters. The results indicate the existence of three health-related behaviour patterns among Brazilian adolescents. The introduction, hypotheses, methods, results and discussion sections are mostly sound and relevant. The references seem relevant. A number, of amendments are needed and they are given either as mandatory ("should") or as optional "may" below.</p> <p>There is no need for professional English language examination.</p> <p>Title The cross-sectional study design should be added to the title. The fact that this is a population-based study may be added to the title.</p> <p>Abstract The participation rate may be added to the abstract.</p> <p>Introduction The introduction is ideally short and relevant.</p> <p>Methods There seems to be different sampling levels in different geographic areas, but all geographic areas include schools and classrooms. The school context is a perfect context for multilevel analyses with individuals nested within classrooms and classrooms nested within schools. Did the authors consider a multilevel design for the analyses? Why did they not conduct such analyses? If they did not consider such a design: why not? The fact that the statistical analyses in this manuscript were conducted based on individuals</p>
-------------------------	--

	<p>only is a major weakness and limitation of manuscript in its present form. The study population was sampled with schools and classrooms as sampling units, but all statistical analyses are solely conducted at the level of individuals. This is surprising, given the fact that health-related behaviours among adolescents have often been shown to be highly clustered within well-defined social contexts such as classrooms and schools. The authors should either explain very carefully in the manuscript why they have made this decision or amend in accordance with the sampling strategy by which the study population was assessed.</p> <p>The response rate is 82.7% which is high but also normal for school studies among adolescents. The high response rate lowers the risk of selection bias, but the risk of selection may still be considered in the manuscript.</p> <p>The questionnaire is based on the Global School-Based student Health Survey, which is good. The authors refer to this survey instead of defining the health-related behaviour variables. The authors should define the variables in the text of the methods section. The only legitimate reason not to do so is if the word number limit of the journal makes it impossible.</p> <p>In the last paragraph on page 9, end of first sentence, the phrasing “parsimonious estimate” does not seem to be professionally statistical.</p> <p>The authors use imputation technique in the statistical analyses. While this technique is currently considered legitimate by many statisticians and epidemiologists, the outcome is still dependent on the assumptions specified concerning the characteristics of those with missing values. The authors may keep the imputations, but these should be complemented with analyses based on restriction, i.e. analyses restricted to only those individuals who with answers on all items including in this particular study. This will be a test of the robustness of the results.</p> <p>The authors should state the reason why linear regression models were chosen. In fact, in the present manuscript this is almost impossible to judge both because the items/variables included are not described (see above) and because assessments of linear relationships between variables are described.</p> <p>Results</p> <p>The results section is relevant. The tables are well defined in terms of time, place and person.</p> <p>Discussion</p> <p>The discussion should be altered in accordance with the amendments which follow from the comments and recommendations stated above.</p> <p>The last paragraph of the discussion section which concerns “health promotion and disease prevention” should be expanded and also specified.</p>
--	---

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name : Helen Sweeting

Institution and Country : MRC/CSO Social & Public Health Sciences Unit,
University of Glasgow, UK

Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below : Patterns of health-related behaviours among Brazilian adolescents. Bmjopen-2016-011571

Comments for the authors.

Thank you for asking me to review this paper which presents an analysis of data on a wide range of health-related behaviours obtained from a sample of over 100,000 Brazilian 9th grade students (83% of those attending school). Gender differences in behaviours are described, factor analysis conducted, resulting in three factors, and the associations between these factors and a number of socio-demographic characteristics are presented.

The paper is generally well written and none of my comments is major; most are matters of detail or grammar.

Abstract

P3, line 22 – suggest ‘... associated WITH each one ...’

P3, line 40 – because the term public school does not mean the same in all countries, suggest that here it says ‘public (compared with private) schools’

Authors: Thank you for these observations. We have amended the abstract accordingly.

Strengths/limitations

Final bullet – does not make sense, as it says ‘a SINGLE, self-reported behaviour’ when this is analysis of multiple behaviours (same issue in the Discussion (p19, line 47)).

Authors: Thank you for this comment. We apologize for the lack of clarity; by “single” we meant only one informant group (students) instead of multiple informant groups, such as teacher, parent and student reports. In order to clarify this, we have revised these sentences as follows:

“Based on the self-reported behaviours of a single group of respondents (the students), which could result in information bias.”

Introduction

P5, line 6 – suggest ‘health-COMPROMISING ...’ (rather than related), since next sentence says these all cause morbidity/mortality, whereas some health-RELATED behaviours (eg exercise) are health promoting.

P5, line 19 – suggest ‘however, some ISSUES remain ...’, since the points listed are not questions.

P5, from line 51 – suggest ‘The factors that shape patterns OF behaviours ... frameworks, MAY VARY BETWEEN high ...

P6, line 7 – suggest cutting second ‘of’ (‘... study how ...’).

Authors: Thank you for these suggestions. We have made all the changes indicated.

Methods

I assume description of the sampling, stratification, etc (p6-7) and analysis (p8-9) are correct and appropriate – my stats isn’t sophisticated enough to be confident on this.

P7, final line – did the questionnaire really ask about occupation/ I thought these were school pupils.

Authors: Thank you for this comment. The questionnaire really asked about occupation, because the survey was carried out with students attending 9th grade (aged between 11 to 19 years old), and in Brazil the Federal Constitution allows adolescents aged 16 or more to have a job, and adolescents

aged 14 or more to have a job such as apprentice or trainee. In this sample, 13% of students self-reported to having a job.

P8, line 23 – refers to Suppl file – should this be Appendix 1?

Authors: Thank you for this suggestion. We have amended this throughout the text.

P8 – categorisation of geographic regions as more/less developed – is this a standard grouping for Brazil? Is there a reference?

Authors: Thank you for this suggestion. In Brazil there are five geographical regions, which have distinct characteristics. One way of aggregating regions is to classify them according to their level of development. We have included the reference used for this classification.

Reference: Steele, EM; Claro, RM, Monteiro, CA. Behavioural patterns of protective and risk factors for non-communicable diseases in Brazil. *Public Health Nutrition*, 2013 17(2), 369–375.

P8-9 – Descriptive analysis and EFA were both conducted according to gender. That's fine, but needs some justification, otherwise some readers might wonder why the authors weren't (also) interested in doing this by age, SES, etc. If gender is a focus, then I think a rationale needs to be provided in the Introduction (this would not be difficult). In addition, having noted separate male and female EFA, the remainder of the analyses does not mention analysis by gender. As a reader, my guess was that this was because there were no significant differences, but it would have been good to know this was correct.

Authors: Thank you for highlighting this issue. We have included the following statement in the introduction regarding gender differences and also clarified that we have not found gender differences in behaviour patterns and therefore the following analyses were not separated by gender, but only adjusted.

Introduction:

"It is also important to assess whether the behaviour patterns vary according to gender, since, historically, trends of behaviours have shown variation by gender and recent research suggests a decrease in gender differences (MacArthur et al., 2012). Findings from such analyses could have important implications for public health interventions to address multiple risk behaviours and its resulting disease burden 5"

Results:

"The clustering of behaviours was similar for males and females; therefore, the following analyses, including the CFA model and the regression models, were performed for the whole sample and adjusted by gender."

Reference:

MacArthur, GJ; Smith, M.C.; Melotti, R.; Heron, J.; Macleod, J.; Hickman, M.; Kipping, R.R.; Campbell, R.; Lewis, G. Patterns of alcohol use and multiple risk behaviour by gender during early and late adolescence: the ALSPAC cohort. *Journal of Public Health* 2012. Vol. 34, No. S1, pp. i20 – i30.

P9, lines 36-40 – is this detail on variance explained needed?

Authors: Thank you for this comment. This information shows readers how we estimated the variance explained by the factors, which is an important component in factor analysis.

P10 – multiple imputation is mentioned. Were analyses conducted on the raw and the imputed datasets? Which results are presented?

Authors: The analyses were conducted on the imputed datasets but also on the raw dataset (for sensitivity analysis). We have only presented the results based on the imputed dataset. We apologize for the lack of clarity in this regard and we have included this statement in the methods section:

“Using WMLSV in Mplus, the missing data were handled as a function of observed covariates, assuming that the missingness mechanism is “missing at random 19.”

“Multiple imputation was performed using the chained equation technique due to the significant proportion of missing values for the mother’s education level (17%, n=18,527), but also included all other variables with a smaller proportion of missing, to create a complete dataset. The distribution of the observed data was used to estimate a set of plausible values for the missing data, including random components in the estimated values 22. We performed analysis using the raw dataset and the imputed dataset (data not shown), for sensitivity analysis. The results from these analyses did not differ. Therefore, the imputed dataset was used to run the linear regression models of factor scores.”

Results

P11, first results para – I think a table providing sample characteristics would have been useful. This could also show numbers with missing data for each of these socio-demographic variables.

Authors: Thank you for this suggestion. We have included a table describing the socio-demographic characteristics of students; however, we disagree with the inclusion of the number of observations for each variable, because we believe this could mislead the readers. Exploratory factor analysis, confirmatory factor analysis and linear regression were all performed on the complete imputed dataset, therefore, if we present a descriptive analysis with missing data, we could confuse the readers. To inform the readers about the magnitude of missing data, we have presented the highest proportion of missing data found in the dataset (mother’s education level) in the text. We hope this is acceptable to the reviewer and editors.

P11, lines 35-37 – males and aggressive behaviour – this is four variables, three dichotomous and one multi-category, so ‘6 percentage points higher’ is unclear.

Authors: Thank you for this observation. We have revised this sentence to clarify the information.

“Males presented a higher proportion of all frequencies of involvement in aggressive behaviour when compared to females (see table appendix 1). Males were also more involved in unsafe sex (twice as often as females) (see table appendix 1).”

P11, lines 44-46 - I do not think ‘the lowest frequencies of protective behaviours represented a small proportion of students’ is correct. One or no days per week for raw veg was reported by 45%, cooked veg 57%, fruit 38% and physical activity 45%.

Authors: We appreciate your comment. We agree with your observation and therefore we have withdrawn this statement.

P12-13 – description of clustering. I think these are factors (first, second, etc), not patterns (they represent an overall pattern). And I wondered slightly about the terms chosen. The second factor is 'health compromising', but the first ('problem behaviours') is also health compromising. Perhaps the second and third could be 'health compromising diet and sedentary behaviours' and 'health promoting diet and physical activity' – that's clearer, even if rather more wordy.

Authors: Thank you for these suggestions. It is usual to use the term "pattern" when referring to factors extracted from factor analysis, because factors represent a latent construct, which can be translated into a pattern. Please, see the citations below for more examples. Regarding the terms chosen to name the patterns, we agree that your suggestion makes the meaning of patterns clearer, and therefore we have replaced the original terms with those you suggested throughout the text.

Reference:

- Newby P. K., Katherine M. S., Tucker L. Empirically Derived Eating Patterns Using Factor or Cluster Analysis: A Review. *Nutrition Reviews*, 2004, 62 (5) 177-203.
- Steele E.M., Claro R.M., Monteiro, C.A. Behavioural patterns of protective and risk factors for non-communicable diseases in Brazil. *Public Health Nutrition*: 17(2), 369–375

P15, Figure 1 – uses 'HealthY PROMOTING' whereas the rest of the paper (I think) uses 'health-enhancing' – I suggest consistent use of terms. It also includes 'unsafe sex', whereas Appendix 1 has this as a heading, not an individual item.

Authors: Thank you for this observation. We have amended the Fig 1 and we have checked throughout the text.

P15, line 12 - 'were weakly correlated' – these (weak) correlations are not shown on Fig 1, so readers do not know how weak they are.

Authors: Thank you for this suggestion. We have included the values for the correlations between the other factors in Fig1, see above.

P16, Table 2 – something has gone a bit wrong with the row alignment on the problem behaviour and maternal education results.

Authors: We have reformatted this table.

Discussion

P17, 3rd para, 1st sentence – unclear.

Authors: Thank you for this comment. We have revised the sentence as follows:

"The health compromising diet and sedentary behaviours pattern was not correlated to the health promoting diet and physical activity pattern (even negatively), indicating that to engage in a risky lifestyle was determined by a different construct than, for example, not engaging in a healthy lifestyle, possible due to different predictors (for example, someone who does no exercise may eat healthily). The health promoting diet and physical activity pattern may be more influenced by family behaviours since their components are commonly shared and promoted by family according to the literature 26.

P18, lines 10-17 – could this sentence be cut? It does not seem to relate to the rest of the paragraph.

Authors: We appreciate your suggestion. The sentence was cut.

P18, line 30 – 'gender differences in behaviours have been decreasing ...' – this is a very general statement – have differences in ALL behaviours decreased?

Authors: Thank you. We have included some examples of behaviours in which the gender gap has been decreasing.

“Other studies have found no gender differences in behaviours, such as alcohol consumption, smoking and drug use, or a decrease in the gap between males and females in Brazil as well as in other parts of the world 29-32.”

P18, line 37 - ‘some similarities and differences were found’ – unclear – similarities and differences between what?

Authors: We explain the similarities and differences in individual characteristics between the three patterns in the following sentences.

“In terms of the characteristics of individuals associated with the three patterns, some similarities and differences were found. Older students (>16 years old) had a higher score in the problem-behaviour patterns than younger students. Since 16 years is older than usual among 9th grade students, it is possible that these students were kept in 9th grade for an extra year due to poor school performance. Older age is also associated with health compromising diet and sedentary behaviours, but the strength of this association was not as marked. Similar associations with age have been reported before and suggest that prevention programs may need to focus more closely on those students possibly struggling with school before they engage in multiple risk behaviours 32.”

P19, line 21 – how is SES conceptualised here? In material terms (not measured in the study) or in terms of educational capital (measured via maternal education)? I think this would be worth an extra sentence.

Authors: Thank you for this suggestion. We have revised the sentence to explain that maternal education may represent educational capital and may also reflect to some extent the material availability, since higher educational level enables people to have jobs that pay more. Moreover, in this survey income was not measured, therefore maternal educational level was the only indicator of SES we had.

“The health compromising diet and sedentary behaviours pattern and health promoting diet and physical activity pattern were associated with higher maternal education. In this study, maternal education was the only socioeconomic status indicator available. The association found is not surprising, since opportunities to engage in physical activities and access to both healthy and unhealthy food seem to depend on socioeconomic status, something that has been described in relation to food acquisition patterns among Brazilians 36.”

P19, line 47 – single behaviour issue, noted earlier.

Authors: Revised.

“Based on the self-reported behaviours of a single group of respondents (the students) which could result in information bias.”

P20, 2nd para – in thinking about interventions, it might also be worth noting prevention activities, which would occur at younger ages, rather than reduction or cessation interventions with older adolescents.

Authors: Thank you for this observation. We have revised the text to include this issue:

“The associations between socio-demographic characteristics and behaviour patterns suggested that older students from more developed regions were the most vulnerable to the health compromising behaviours patterns. Older students in 9th grade are probably those experiencing difficulties at school and have been left behind. Therefore interventions could target this group to tackle multiple health risk behaviours. On the other hand, preventive strategies should be directed to students at an early age.”

Supplementary files 2 and 3 – I would have found an extra column, with significance of the gender difference, useful, because these are quite large tables and it is difficult to assess quickly where the main differences are. A further column could present numbers with missing data for each behavioural variable.

Authors: Thank you for this suggestion. Regarding the significance of gender difference, these two tables are descriptive ones, and therefore, not adjusted for potential confounders. Thus, we believe that providing significance in this table could be misleading to the reader, since the associations found would be crude. We prefer let the reader assess the differences in percentage points and take his own conclusions on the magnitude of gender differences. Regarding the missing values, we have answered this issue in your comment about the results section.

Reviewer: 2

Reviewer Name: Martin Lindström

Institution and Country: Department of Clinical Sciences in Malmö Lund University Sweden

Please state any competing interests or state 'None declared': None declared.

Please leave your comments for the authors below

Manuscript: "Patterns of health-related behaviours among Brazilian adolescents"

The cross-sectional study from Brazil investigates the clustering of multiple health-related behaviours among adolescents and socio-demographic characteristics associated with these clusters. The results indicate the existence of three health-related behaviour patterns among Brazilian adolescents. The introduction, hypotheses, methods, results and discussion sections are mostly sound and relevant. The references seem relevant. A number, of amendments are needed and they are given either as mandatory ("should") or as optional "may" below.

There is no need for professional English language examination.

Title: The cross-sectional study design should be added to the title. The fact that this is a population-based study may be added to the title.

Author: Thank you for the suggestion. We have revised the title to: "Patterns of health-related behaviours among adolescents: a cross-sectional study based on the National Survey of School Health Brazil 2012".

Abstract: The participation rate may be added to the abstract.

Author: Thank you for this suggestion. We have included this information in the abstract.

"Participants: 104,109 Brazilian 9th Grade Students from public and private schools (response rate = 82.7%)."

Introduction

The introduction is ideally short and relevant.

Author: Thank you for this comment.

Methods

There seems to be different sampling levels in different geographic areas, but all geographic areas include schools and classrooms. The school context is a perfect context for multilevel analyses with individuals nested within classrooms and classrooms nested within schools. Did the authors consider a multilevel design for the analyses? Why did they not conduct such analyses? If they did not consider such a design: why not? The fact that the statistical analyses in this manuscript were conducted based on individuals only is a major weakness and limitation of manuscript in its present form. The study population was sampled with schools and classrooms as sampling units, but all statistical analyses are solely conducted at the level of individuals. This is surprising, given the fact that health-related behaviours among adolescents have often been shown to be highly clustered within well-defined social contexts such as classrooms and schools. The authors should either explain very carefully in the manuscript why they have made this decision or amend in accordance with the sampling strategy by which the study population was assessed.

Authors: Thank you for highlighting this issue. We acknowledge that having a dataset with students nested in schools, the multilevel analysis should be considered. An issue regarding multilevel analysis when using survey datasets is that it is not possible to consider sample weights in the analysis. On the other hand, when using survey data, the Stata command survey (SVY) takes into consideration all the sample structure (strata, primary, secondary and tertiary sample units and sample weights). We did not have a contextual hypothesis to test using multilevel analysis; therefore, the indication would be based only in the possible lack of independence of data, which somehow the SVY command also takes account of. We did perform multilevel analysis, with the limitations already outlined, to test for clustering in our outcomes and we found a small percentage of the variance explained at school level. When we compared the coefficients obtained from multilevel linear regression analysis with those obtained by regular linear regression analysis (with the svy prefix) we found very similar results in terms of both magnitude and statistical significance.

We have included the following statement in the discussion:

“We also acknowledge that due to the nature of the dataset (students nested within schools), the observations are not independent and, therefore, multilevel analysis could be a valuable alternative when describing the pattern scores. However, multilevel analysis using STATA does not fit sample weights used in the sampling procedure, which could bias the results. On the other hand, the SVY prefix on Stata software, which we have used in multiple linear regression takes into consideration the sample structure (strata, sample units and sample weights) when calculating the estimates. We chose to use multiple linear regression with the SVY prefix instead of multilevel analysis, both because the clustering effect we found was small and we did not have a contextual hypothesis to test.”

The response rate is 82.7% which is high but also normal for school studies among adolescents. The high response rate lowers the risk of selection bias, but the risk of selection may still be considered in the manuscript.

Authors: We agree that despite the high response rate it is still possible to have selection bias. Therefore we have included this statement in the discussion section, as follows:

“In addition, despite the high response rate obtained, which reduces the possibility of selection bias, it is still possible that students that had refused to participate or were not at school during the data collection were also at higher risk for the factors assessed.”

The questionnaire is based on the Global School-Based student Health Survey, which is good. The authors refer to this survey instead of defining the health-related behaviour variables. The authors

should define the variables in the text of the methods section. The only legitimate reason not to do so is if the word number limit of the journal makes it impossible.

Authors: The Appendix 1 of our paper is a description of all the health behaviors and health indicators used in our study, including the variable, the question and the options of answer (operationalization). In the methods section we refer to this appendix as follows:

“Health-related behaviours analyzed in the study included involvement in physical fights, fights with guns or other weapons (knives, bottles, etc), bullying behaviour (aggressor/bully), alcohol use, drug use, smoking, sexual behaviour, physical activity and sedentary behaviour, and dietary intake of healthy and unhealthy food indicators (Appendix 1).”

In the last paragraph on page 9, end of first sentence, the phrasing “parsimonious estimate” does not seem to be professionally statistical.

Authors: We have revised this expression.

“After the factor solution was determined, Confirmatory Factor Analysis (CFA) was performed for the factors retained in EFA and the observed variables with loading above 0.3 as indicators of each factor, in order to reduce the complexity of the model.”

The authors use imputation technique in the statistical analyses. While this technique is currently considered legitimate by many statisticians and epidemiologists, the outcome is still dependent on the assumptions specified concerning the characteristics of those with missing values. The authors may keep the imputations, but these should be complemented with analyses based on restriction, i.e. analyses restricted to only those individuals who with answers on all items including in this particular study. This will be a test of the robustness of the results.

Authors: We appreciate your suggestion. We have performed analysis on the raw dataset (only complete answers) and compared to the results from the imputed dataset. We found no remarkable differences, only small changes of second decimal place. We therefore included a statement about this sensitivity analysis in the methods section as follows:

“Multiple imputation was performed using the chained equation technique due to the significant proportion of missing values for the mother’s education level (17%, n=18,527), but also included all other variables with a smaller proportion of missing, to create a complete dataset. The distribution of the observed data was used to estimate a set of plausible values for the missing data, including random components into the estimated values. We performed analysis using the raw dataset and the imputed dataset (data not shown), for sensitivity analysis. The results from these analyses did not differ. Therefore, the imputed dataset was used to run the linear regression models of factor scores.”

The authors should state the reason why linear regression models were chosen. In fact, in the present manuscript this is almost impossible to judge both because the items/variables included are not described (see above) and because assessments of linear relationships between variables are described.

Authors: We have used linear regression because the factor scores estimated for each factor, for each student are continuous variables, and we tested the variables for the linear regression assumptions. We apologize for not being clearer about this specific point. The text was amended as follows:

“Following CFA, factor scores (continuous variable) were estimated, through regression analysis, for

each one of the factors, to identify an individual's placement within the factors 21. Afterwards, we performed linear regression models to describe sociodemographic characteristics associated with each pattern of behaviour (factor scores) in multiple adjusted models. The linear regression assumptions were met.”

Results

The results section is relevant. The tables are well defined in terms of time, place and person.

Discussion

The discussion should be altered in accordance with the amendments which follow from the comments and recommendations stated above.

Authors: Thank you for the comments. We have included all the necessary amendments following the previous comments.

The last paragraph of the discussion section which concerns “health promotion and disease prevention” should be expanded and also specified.

Authors: We have included some extra-sentences to the final paragraph, as follows:

“Our results show that health behaviours tend to cluster among Brazilian adolescents, with these eighteen behaviours grouping into three patterns. Interventions regarding health promotion and disease prevention should be designed focusing on behaviour patterns instead of single behaviours, as it is often the case. High risk behaviours tend to cluster in the same individuals and it seems rather inefficient to design programmes addressing single unhealthy behaviours. The associations between socio-demographic characteristics and behaviour patterns suggested that older students from more developed regions were the most vulnerable to the health compromising behaviours patterns. Older students in 9th grade are probably those experiencing difficulties at school and have been left behind. Therefore interventions could target this group to tackle multiple health risk behaviours. On the other hand, preventive strategies should be directed to students at an early age. Future research should also take a step further in this field in trying to understand the mechanisms that give rise to health behaviour clustering, together with their implications for interventions 3”

VERSION 2 – REVIEW

REVIEWER	Helen Sweeting MRC/CSO Social & Public Health Sciences Unit, University of Glasgow, UK
REVIEW RETURNED	18-Aug-2016

GENERAL COMMENTS	<p>Patterns of health-related behaviours among adolescents: a cross-sectional study based on the National Survey of School Health Brazil 2012</p> <p>Thank you for sending me this revised manuscript – the authors have done a good job in addressing the requested amendments. I only have a very few, very small comments (mainly typos) this time round which I am sure the authors could easily deal with, without me having to check again.</p> <p>P6, line 21 – ‘suggests a decrease in gender differences’ – think this is too general and suggest ‘in gender differences IN CERTAIN</p>
-------------------------	--

	<p>BEHAVIOURS’.</p> <p>P8 – the authors note they have added a reference (Steele), but I can’t see this in the manuscript.</p> <p>P10, line 57 – should this read ‘... but WAS USED in all other analySES’?</p> <p>P13, lines 16-23 – suggest these are shifted to after the ref to table 2 (current line 34), with slight amendments: ‘HOWEVER the clustering of behaviours was ... the following analySES, including the CFA model AND the regression ...’</p> <p>P18, line 14 – suggest cutting ‘in the compromising behaviour, which included’.</p> <p>P18, line 24 – please confirm whether the Netherlands study was of adolescents.</p> <p>P19, lines 7-12 – suggest ‘... findings suggest that ENCOURAGING health promoting ...[cut pattern] may not affect the adoption [cut of the] health compromising ...’.</p> <p>P19, line 21 – suggest ‘gender differences in RATES OF behaviours’.</p> <p>P20, line 23 – suggest cutting ‘it’ (Among these is the).</p>
--	--

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name

Helen Sweeting

Institution and Country

MRC/CSO Social & Public Health Sciences Unit,

University of Glasgow, UK

Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below

Patterns of health-related behaviours among adolescents: a cross-sectional study based on the National Survey of School Health Brazil 2012

Thank you for sending me this revised manuscript – the authors have done a good job in addressing the requested amendments. I only have a very few, very small comments (mainly typos) this time round which I am sure the authors could easily deal with, without me having to check again.

Authors: We appreciate your comments.

P6, line 21 – ‘suggests a decrease in gender differences’ – think this is too general and suggest ‘in gender differences IN CERTAIN BEHAVIOURS’.

Authors: Thank you for this suggestion. We have included “in certain behaviours” in this phrase.

P8 – the authors note they have added a reference (Steele), but I can’t see this in the manuscript.

Authors: Thank you for this observation. We have included this reference in the paper (reference number 8).

P10, line 57 – should this read ‘... but WAS USED in all other analySES’?

Authors: Yes. We have ammended this phrase as you suggested.

P13, lines 16-23 – suggest these are shifted to after the ref to table 2 (current line 34), with slight amendments: ‘HOWEVER the clustering of behaviours was ... the following analySES, including the CFA model AND the regression ...’

Authors: We apologize for spelling errors, and have thoroughly reviewed the manuscript.

P18, line 14 – suggest cutting ‘in the compromising behaviour, which included’.

Authors: Suggestion accepted.

P18, line 24 – please confirm whether the Netherlands study was of adolescents.

Authors: This study was among adolescents and adults, however the factor analyses were conducted per age group. We have included the information that we were mentioning the adolescents’ results, for clarity purpose.

“In the Netherlands, researchers also found two patterns of risky behaviours among adolescents. These patterns differed from ours regarding the variables that clustered together. Alcohol consumption, unsafe sex and vigorous physical activity clustered together in one pattern, and violent behaviour, smoking and drug use in another¹⁰.”

P19, lines 7-12 – suggest ‘... findings suggest that ENCOURAGING health promoting ...[cut pattern] may not affect the adoption [cut of the] health compromising ...’.

Authors: Thank you for this suggestion. We have ammended this sentence.

P19, line 21 – suggest ‘gender differences in RATES OF behaviours’.

Authors: Thank you for this suggestion. We have ammended this phrase.

P20, line 23 – suggest cutting ‘it’ (Among these is the).

Authors: Thank you for this suggestion. We have cut “it” in this sentence..