

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 multiple lifestyle risk factors and their link to mental health in the German adult population: a cross-sectional study
AUTHORS	Atzendorf, Josefine; Apfelbacher, Christian; Gomes de Matos, Elena; Kraus, L; Piontek, D

VERSION 1 – REVIEW

REVIEWER	Sabine Rohrmann University of Zurich, Switzerland
REVIEW RETURNED	18-Feb-2018

GENERAL COMMENTS	This is an interesting manuscript looking at lifestyle and health in a German population. I have several comments mainly on the Methods and Results. I am not familiar with Latent Class Analysis and would like the manuscript to be reviewed by a statistician. Methods: - For readers who do not know ESA: how were participants recruited and who conducted that survey? Eligible were all residents in Germany 18-64 years old? - Just for clarification: risky tobacco consumption is daily smoking independent of number of cigarettes? EHD is at least one heavy drinking event within the last 30 days? "non-prescribed analgesics" also include, for example, aspirin? What about "mis"-use of prescribed drugs? In the Discussion, the authors state that "the ESA 2015 was unfortunately not able to distinguish between the usage of prescribed and non-prescribed opioid and non-opioid pharmaceuticals within the last 30 days". This sounds like a contradiction to what is described in the Methods section (clearly stating "non-prescribed analgesics"). - Has the validation of the FFQ used in this study been published or details of the FFQ in general. If yes, please give a reference or provide some detail here. However, it is rather superficial to base the definition of a healthy diet on only five food items. Foods such as meat and vegetables are apparently not included. - Physical activity recommendations by WHO (http://www.who.int/dietphysicalactivity/leaflet-physical-activity-recommendations.pdf?ua=1): "Adults aged 18–64 should accumulate at least 150 minutes of moderate-intensity aerobic physical activity throughout the week or do at least 75 minutes of vigorous-intensity aerobic physical activity throughout the week or an equivalent combination of moderate- and vigorous-intensity activity. " It seems to me that the authors' definition is based only on vigorous physical activity and thus requires twice the amount that the WHO definition. Results:
-------------------------	--

	- For the reader, it would be informative to see a baseline description of the study population. Discussion: - The authors should discuss whether the short FFQ used is appropriate to assess the "healthiness" of a person's diet. The same is true of physical activity assessment, for which usually validated questionnaires are used.
--	---

REVIEWER	Lucia Valmaggia Kin's College London, Institute of Psychiatry, Psychology and Neuroscience, London, UK
REVIEW RETURNED	07-Mar-2018

GENERAL COMMENTS	This is a very interesting and well written manuscript. The study has a large sample size, the methodology is well described and the results are interesting. The discussion critically reviews the results.
-------------------------	---

REVIEWER	Prof Andrew McVicar Anglia Ruskin University, UK
REVIEW RETURNED	01-May-2018

GENERAL COMMENTS	This study has extracted data from a national survey (ESA15) to explore multiple lifestyle factors for risk clusters identified from those data. The study is well written and presented. I am not familiar with the range of Goodness-of-fit measures but the statistics appear to have been applied diligently and outcomes are described well. It is claimed that this is the first such analysis to be conducted in Germany. Some related studies are mentioned including systematic reviews which appear to endorse this position regarding multiple lifestyle analysis. If there is a weakness then for me it is the limitations set by the ESA survey data. There are a number of caveats regarding deficits in those data that are highlighted by the authors, primarily in the Discussion. I would like to have seen within the Methods section an overview of what the ESA does or does not provide in terms of enabling a sufficient depth and breadth of data to be extracted and utilised for this study, in particular concerning the rationale for interpretative contexts for two of the main Findings: 1. The ESA survey of substance abuse makes a distinction between cannabis users and users of a range of other illicit drugs. Cannabis was clearly the main substance within these categories but the sample Class for this study was clustered around 'Polysubstance use lifestyle'. Table 1 and Figure 1 distinguish between the two categories and the reason for this mergence is unclear - they relate in terms of behaviour but not in terms of drug class. What was the rationale for this merging? 2. Figure 1 presents an interesting depiction of response probabilities. It is notable that all Classes score highly on 'unhealthy diet' and 'low physical activity' presumably as a consequence of the scope of the ESA. For comparative purposes Class 1 was referenced in this study as 'healthy'. Since all Classes show this pattern then making comparisons between the Classes is valid but I would like to have seen some recognition/discussion of this restriction since of course poor diet and low activity are both significant health concerns and currently are an important public health focus. The Methods section identifies a diet assessment and
-------------------------	--

	also a rating of activity as 30mins/day on 5days/week - did the ESA study recruit largely to those lifestyle classes? If so are they regarded by the ESA report as 'poor'? Some further points for clarification: 3. Please clarify that the paper/pencil, telephone and internet-based collection methods, together with other details presented in the Methods section, relate to the construction of the ESA database and not to additional activities performed for this study. 4. Following on from 3, above, what evidence is there for the quality of the ESA data? Comments early in the Methods as to that quality would be beneficial. For example, the potential for respondent bias is noted in the Discussion. Additionally, I was intrigued to see that diet assessment included the frequency of usage of herbs since these are almost ubiquitous in prepared food - sandwiches, salads, cooked foods - and so are difficult to avoid! 5. Table 4 highlights a range of mental health problems but in the text (i.e. Line 45, page 4; and Line 21 page 6) these are shown as 'i.e. depression' when I think this should read 'e.g. depression'. 6. Line 15, page 12. It is noted that respondents in the Noble et al. study belonged to a 'healthy' class. In view of 2, above, this requires clarifying in the present context where 'healthy' may be defined differently.
--	--

REVIEWER	Rob Saunders UCL, UK
REVIEW RETURNED	18-May-2018

GENERAL COMMENTS	This study investigated sub-groups of risk taking behaviour in a German population, using latent class analysis. Four latent classes were identified, and predictors of class membership were further explored in the study. The paper benefits from a substantial dataset with regard to both sample numbers and available variables for exploration. I would however recommend some minor improvements to the manuscript to increase its reach and interpretability. These are mainly about clarifying certain areas of your analysis for the reader, who for this journal will include a range of knowledge and specialities. These are listed below: 1. Abstract: "A total of four classes were extracted" however you have not discussed that latent class analysis was performed, and therefore the reader may be unaware what you mean by 'class'. I recommend this is made clearer in the abstract. 2. Similar to the above, at the end of the introduction it is stated that the aim is to use "extracted health, lifestyle classes.." (page 5) but you have not introduced what a 'class' is. In the introduction you refer to "clusters" and therefore more consistency with regard to the terms 'clusters' and 'classes' would help clarify. Introducing latent class as an approach to identifying classes in the introduction may help clarify further. 3. Please provide citations for your decision making around the number of classes you selected. The sentence "A low BIC, high entropy..." although non-controversial this should be back-up by some literature. This is mainly to help readers understand why your decisions were made, and to guide researchers towards best practice. 4. Please define what you mean by "complex mixture term" as isn't clear what it is or why you used from the reader's perspective (end of page 6). 5. From reading the next section "The conditional class membership
-------------------------	--

	probabilities were further used”, I believe you assigned each case to the class to which they had the highest probability of profile membership and then ran multinomial regression analysis. I would make this clearer as readers who are not knowledgeable about LCA may not understand this part. If you didn’t do this then please explain further. 6. Please explain to the reader why you included a “complex mixture term to adjust standard errors” (page 7, first paragraph). It won’t be obvious to all readers 7. Please clarify in the same paragraph whether your multinomial regression analysis included all of your predictor variables (one model) or whether you entered each predictor univariately. From the notes in the table it is implied that you included together but please specify in this section of the method. Having split tables makes them easier to read, but also confuses the situation. 8. Class 3 – if higher scores on the ‘physical health’ item indicate poorer health, then are this class not more likely to report poor health?
--	---

REVIEWER	Alan Watkins Swansea University, UK
REVIEW RETURNED	31-May-2018

GENERAL COMMENTS	Summary: from the (necessarily) limited information provided in the manuscript, it would seem that the statistical analysis of the data in this study is appropriate, and that it has been competently performed - however, much more clarity is required on almost all statistical aspects of the manuscript, starting from a definition of latent class analysis (LCA) or, at least, a working description of its purpose – to identify an unknown (a priori) number of groups in a sample. The matter of what an LCA adds to one’s understanding also needs to be addressed more thoroughly. Specific points requiring clarification in the manuscript and/or comment from the authors: 1. The role of weighting in the ESA sample needs to be clarified; the stated numbers and percentages in the “Sampling & procedure” paragraph need checking. 2. Given that LCA depends heavily (entirely?) on the risk factors, there needs to be better justification of those used here – if three (of eight) factors relate to cannabis, illicit drugs, and non-prescribed analgesics, it is hardly surprising that one latent class (albeit not identified by every modelling exercise) may be labelled “polysubstance use”. 3. Some of the definitions used to dichotomize risk factors seem arbitrary (as the authors admit); some result in thresholds that seem severe compared with others; and at least one needs checking – if I understand correctly, it is sufficient to undertake physical activity for less than 30 minutes per day on three days a week to be regarded as at risk as physically inactive, which is subtly different from what is written on page 6. This definition seems to set a very high threshold. 4. Some sociodemographic factors seem to be very traditional – I was surprised that status did not explicitly account for “cohabiting” (conflate with “married”?) or “separated” (conflate with “divorce”?), as would almost certainly be done in the UK! 5. The paragraph on “Data analysis” on page 6 needs thorough revision, starting with a reference (such as Hagenaars & McCutcheon, 2002) for LCA. The method of estimation (EM algorithm?) needs to be mentioned, while discussion of starting values, iterations, convergence can either be omitted entirely, quickly summarised but without any technical terms (probably best
-------------------------	---

	for the BMJ), or further details should be provided – at the moment, the account falls between these stools. More details on the “complex mixture term” are definitely required! 6. The second paragraph needs a reference to cover the multinomial logistic regression modelling; the reference to STATA needs to appear in the bibliography (cf the previous paragraph and Mplus). 7. Discussion of Table 1 needs to cover the use of weighting factors, and to compare weighted and unweighted percentages. This should be linked with the enhanced description of the ESA sample, previously mentioned. 8. The reference to “bivariate residual statistics” needs clarification (may one assume that these are “standardised residuals”?), or, perhaps better here, a more reader-friendly phrasing, such as “Residual-based goodness-of-fit showed only minor violations ...” 9. Table 2 could omit the first model (with only one class), and show subsequent changes in BIC relative to that model. 10. More discussion need on the emerging classes, and the resulting partition of the ESA sample – of particular interest are those cannabis/illicit drug/non-prescribed analgesic users NOT categorised in class 4. Finally, the standard of written English is generally good; in Tables 3 & 4 there is “und” rather than “and”.
--	--

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: Sabine Rohrmann

Institution and Country: University of Zurich, Switzerland Competing Interests: None declared.

This is an interesting manuscript looking at lifestyle and health in a German population. I have several comments mainly on the Methods and Results. I am not familiar with Latent Class Analysis and would like the manuscript to be reviewed by a statistician.

Methods: - For readers who do not know ESA: how were participants recruited and who conducted that survey? Eligible were all residents in Germany 18-64 years old?

Response: More information about the recruiting of the participants was added. The participants had to be 18 to 64 years old to take part in the survey. See p. 5, l. 7-15:

“The ESA is a population-representative cross-sectional survey assessing substance use among the general population in Germany [21] and has been conducted at regular intervals since the 1980s by the IFT Institut für Therapieforschung in Munich, Germany. The adjusted sample (weighted for region, age, gender and education) comprised 9,204 persons aged 18 to 64 years. The sample was drawn by means of a random two-stage selection procedure. At the first stage, 254 sample points (cities, communities) were drawn. Afterwards, the target population was randomly chosen from the sample points’ population registers. Paper-and-pencil, telephone and internet-based methods were used to collect the data (net response rate = 52.2%). Participation was voluntary and could be terminated at any time. All data was used strictly confidentially and anonymously.”

- Just for clarification: risky tobacco consumption is daily smoking independent of number of cigarettes? EHD is at least one heavy drinking event within the last 30 days?

Response: The reviewer understood everything correctly. The author rephrased the sentences to clarify the structure of the selected variables for the latent class analysis.

"Risky tobacco consumption was defined as daily smoking of at least one cigarette, cigar, pipe or cigarillo within the last 30 days." (p.5, l.28-29)

"The daily consumption of 12 grams (women) or 24 grams (men) of pure alcohol within the last 30 days was characterised as risky alcohol consumption. Episodic heavy drinking (EHD) was defined as consumption of five or more drinks of alcohol consumed on a single day for at least one time within the last 30 days." (p.5, l.30-33)

"Non-prescribed analgesics" also include, for example, aspirin? What about "mis"-use of prescribed drugs? In the Discussion, the authors state that "the ESA 2015 was unfortunately not able to distinguish between the usage of prescribed and non-prescribed opioid and non-opioid pharmaceuticals within the last 30 days". This sounds like a contradiction to what is described in the Methods section (clearly stating "non-prescribed analgesics").

Response: Yes, the term "non-prescribed analgesics" includes both, non-prescribed opioid and non-opioid pharmaceuticals. These include for example Aspirin, Ibuprofen, Codein and more.

Unfortunately, the limitation was stated incorrectly in the discussion. The ESA 2015 is not able to distinguish between weekly consumption of non-prescribed opioid analgesics and weekly consumption of non-prescribed non-opioid analgesics. The incorrect statement has been corrected.

Further, the ESA 2015 is not able to assess if participants would misuse prescribed pharmaceuticals. A statement in this regard was added in the discussion.

"Unfortunately, the ESA 2015 data collection did not distinguish between the usage of non-prescribed opioid and non-opioid pharmaceuticals. Furthermore, it was not assessed if participants had misused prescribed pharmaceuticals." (p. 16, l. 11-14)

- Has the validation of the FFQ used in this study been published or details of the FFQ in general. If yes, please give a reference or provide some detail here. However, it is rather superficial to base the definition of a healthy diet on only five food items. Foods such as meat and vegetables are apparently not included.

Response: The validity and reliability of the food frequency questionnaire are confirmed. Results of the dissertation, which contains validation data, are published online: <http://archiv.ub.uni-marburg.de/diss/z1998/0303/html/frame.htm> The URL has been added in the reference list and limitations of the food frequency questionnaire are now described in the discussion section.

"Furthermore, nutrition was assessed with a food frequency questionnaire (LML-6) consisting of six items due to limited space [22]. Even though the LML-6 was validated [22], a different questionnaire with more items could have assessed nutrition and unhealthy diet more precisely." (p. 14, l. 10-13)

- Physical activity recommendations by WHO (<http://www.who.int/dietphysicalactivity/leaflet-physical-activity-recommendations.pdf?ua=1>): "Adults aged 18–64 should accumulate at least 150 minutes of moderate-intensity aerobic physical activity throughout the week or do at least 75 minutes of vigorous-intensity aerobic physical activity throughout the week or an equivalent combination of moderate- and vigorous-intensity activity. " It seems to me that the authors' definition is based only on vigorous physical activity and thus requires twice the amount that the WHO definition.

Response: The American College of Sports Medicine and the American Heart Association recommended moderate-intense activity for a minimum of 30 min on five days each week.

The questions for assessing physical activity have also been used in the German Health Interview and Examination Survey for Adults (DEGS1) study. According to Krug et al. the questions describe moderate-intense physical activity.

Reference: Krug, S., Jordan, S., Mensink, G. et al.

Physical activity Results of the German Health Interview and Examination Survey for Adults (DEGS1). Bundesgesundheitsbl. (2013) 56: 765.

<https://doi.org/10.1007/s00103-012-1661-6>

https://static-content.springer.com/esm/art%3A10.1007%2Fs00103-012-1661-6/MediaObjects/103_2012_1661_MO1_ESM.pdf

Unfortunately, the ESA 2015 did not assess vigorous-intensity activity. Therefore, a statement was added in the discussion, that the amount of physical activity might have been underestimated in the sample. As a consequence, the questions for assessing physical activity have been completely revised for the currently ongoing ESA 2018.

“In addition, the ESA 2015 was only able to assess moderate-intensive physical activity and not vigorous-intensive physical activity, which could have led to an underestimation of physical activity in the sample. As a consequence, the questions for assessing physical activity have been completely revised for future assessments in ESA.” (p. 14, l.7-10)

Results: - For the reader, it would be informative to see a baseline description of the study population.

Response: A baseline description of the study population is provided in the section “Sample and procedure”.

“Overall, 5090 women (49.6%) and 4114 men (50.4%) with an average age of 38.3 years (SD = 14.7) participated in this study.” (p.5, l.16-17)

Discussion: - The authors should discuss whether the short FFQ used is appropriate to assess the "healthiness" of a person's diet. The same is true of physical activity assessment, for which usually validated questionnaires are used.

Response: A statement on the assessment of physical activity and nutrition and their weaknesses has been added in the method section as well as in the discussion.

“Furthermore, nutrition was assessed with a food frequency questionnaire (LML-6) consisting of six items due to limited space [22]. Even though the LML-6 was validated [22], a different questionnaire with more items could have assessed nutrition and unhealthy diet more precisely.” (p. 14, l. 10-13)

“In addition, the ESA 2015 was only able to assess moderate-intensive physical activity and not vigorous-intensive physical activity, which could have led to an underestimation of physical activity in the sample. As a consequence, the questions for assessing physical activity have been completely revised for future assessments in ESA.” (p. 14, l.7-10)

Reviewer: 2

Reviewer Name: Lucia Valmaggia

Institution and Country: King's College London, Institute of Psychiatry, Psychology and Neuroscience, London, UK Competing Interests: None declared

This is a very interesting and well written manuscript.

The study has a large sample size, the methodology is well described and the results are interesting. The discussion critically reviews the results.

Response: Thank you very much!

Reviewer: 3

Reviewer Name: Prof Andrew McVicar

Institution and Country: Anglia Ruskin University, UK Competing Interests: None declared

This study has extracted data from a national survey (ESA15) to explore multiple lifestyle factors for risk clusters identified from those data. The study is well written and presented. I am not familiar with the range of Goodness-of-fit measures but the statistics appear to have been applied diligently and outcomes are described well.

It is claimed that this is the first such analysis to be conducted in Germany. Some related studies are mentioned including systematic reviews which appear to endorse this position regarding multiple lifestyle analysis.

If there is a weakness then for me it is the limitations set by the ESA survey data. There are a number of caveats regarding deficits in those data that are highlighted by the authors, primarily in the Discussion. I would like to have seen within the Methods section an overview of what the ESA does or does not provide in terms of enabling a sufficient depth and breadth of data to be extracted and utilised for this study, in particular concerning the rationale for interpretative contexts for two of the main Findings:

1. The ESA survey of substance abuse makes a distinction between cannabis users and users of a range of other illicit drugs. Cannabis was clearly the main substance within these categories but the sample Class for this study was clustered around 'Polysubstance use lifestyle'. Table 1 and Figure 1 distinguish between the two categories and the reason for this merging is unclear - they relate in terms of behaviour but not in terms of drug class. What was the rationale for this merging?

Response: In order to address the issue and in accordance with reviewer 5 we have changed the label of class 4 to „Cumulate risk factors lifestyle“

2. Figure 1 presents an interesting depiction of response probabilities. It is notable that all Classes score highly on 'unhealthy diet' and 'low physical activity' presumably as a consequence of the scope of the ESA. For comparative purposes Class 1 was referenced in this study as 'healthy'. Since all Classes show this pattern then making comparisons between the Classes is valid but I would like to have seen some recognition/discussion of this restriction since of course poor diet and low activity are both significant health concerns and currently are an important public health focus. The Methods section identifies a diet assessment and also a rating of activity as 30mins/day on 5days/week - did the ESA study recruit largely to those lifestyle classes? If so are they regarded by the ESA report as 'poor'?

Response: The ESA is a representative survey for the general adult population in Germany. The calculation of the four lifestyle risk factor classes is not related to the scope or recruiting strategy of the ESA. More information about the recruiting of the participants was added.

“The ESA is a population-representative cross-sectional survey assessing substance use among the general population in Germany [21] and has been conducted at regular intervals since the 1980s by the IFT Institut für Therapieforchung in Munich, Germany. The adjusted sample (weighted for region, age, gender and education) comprised 9,204 persons aged 18

to 64 years. The sample was drawn by means of a random two-stage selection procedure. At the first stage, 254 sample points (cities, communities) were drawn. Afterwards, the target population was randomly chosen from the sample points' population registers. Paper-and-pencil, telephone and internet-based methods were used to collect the data (net response rate = 52.2%). Participation was voluntary and could be terminated at any time. All data was used strictly confidentially and anonymously." (p. 5, l. 7-15)

The results of poor diet and poor physical activity in all classes are discussed in detail:

"Although the results displayed that more than half of the German population practiced a relatively healthy lifestyle regarding consumption of alcohol, tobacco or illicit drugs, even individuals in the healthy lifestyle class show low physical activity and unhealthy diet." (p.13, l. 31-35).

And one conclusion states: "3) The high probabilities for unhealthy diet and low physical activity emphasize the importance of the promotion of regular physical activity and healthy nutrition in wide sections of the German population through intervention and prevention measures." (p.17, l.16-22)

If we have not answered the comment by the reviewer sufficiently, we kindly ask for further guidance.

Some further points for clarification:

3. Please clarify that the paper/pencil, telephone and internet-based collection methods, together with other details presented in the Methods section, relate to the construction of the ESA database and not to additional activities performed for this study.

Response: Yes, the mention of the modes of administration refers to data collection and not to additional activities performed for this study.

4. Following on from 3, above, what evidence is there for the quality of the ESA data? Comments early in the Methods as to that quality would be beneficial. For example, the potential for respondent bias is noted in the Discussion. Additionally, I was intrigued to see that diet assessment included the frequency of usage of herbs since these are almost ubiquitous in prepared food - sandwiches, salads, cooked foods - and so are difficult to avoid!

Response: The strengths of the survey were added to the discussion, since it seems more naturally to discuss limitations and strengths in the same section.

"Despite these limitations, the ESA provides a large sample and a good response rate [21]. Furthermore, based on the use of different modes of administration and weighting procedures, the ESA provides representative estimates for lifestyle risk factors in the general adult population in Germany." (p.16, l.19-22)

The limitations of the food frequency questionnaire are now mentioned at the end of the discussion. The question about herbs indicates the consumption of fresh herbs and not herbs in prepared food. The term "fresh herbs" has been added in the method section (p.6, l.18).

"Furthermore, nutrition was assessed with a food frequency questionnaire (LML-6) consisting of six items due to limited space [22]. Even though the LML-6 was validated [22], a different questionnaire with more items could have assessed nutrition and unhealthy diet more precisely." (p. 14, l. 10-13)

5. Table 4 highlights a range of mental health problems but in the text (i.e. Line 45, page 4; and Line 21 page 6) these are shown as 'i.e. depression' when I think this should read 'e.g. depression'.

Response: The sentence has been corrected considering the suggestion.

6. Line 15, page 12. It is noted that respondents in the Noble et al. study belonged to a 'healthy' class. In view of 2, above, this requires clarifying in the present context where 'healthy' may be defined differently.

Response: The sentence has been modified.

"In a systematic review of multiple lifestyle risk factors, Noble and colleagues [7] reported that the majority of the respondents (81 %) in the included studies belonged to a class, which did not show any lifestyle risk factors." (p.13, l.7-9)

Reviewer: 4

Reviewer Name: Rob Saunders

Institution and Country: UCL, UK

Competing Interests: None declared

This study investigated sub-groups of risk taking behaviour in a German population, using latent class analysis. Four latent classes were identified, and predictors of class membership were further explored in the study. The paper benefits from a substantial dataset with regard to both sample numbers and available variables for exploration. I would however recommend some minor improvements to the manuscript to increase its reach and interpretability. These are mainly about clarifying certain areas of your analysis for the reader, who for this journal will include a range of knowledge and specialties. These are listed below:

1. Abstract: "A total of four classes were extracted" however you have not discussed that latent class analysis was performed, and therefore the reader may be unaware what you mean by 'class'. I recommend this is made clearer in the abstract.

Response: The authors revised the abstract and added more information about the methods used.

"Results: A latent class analysis was applied to identify patterns of lifestyle risk factors and a multinomial logistic regression was carried out to examine associations between the extracted classes and external factors."

2. Similar to the above, at the end of the introduction it is stated that the aim is to use "extracted health, lifestyle classes.." (page 5) but you have not introduced what a 'class' is. In the introduction you refer to "clusters" and therefore more consistency with regard to the terms 'clusters' and 'classes' would help clarify. Introducing latent class as an approach to identifying classes in the introduction may help clarify further.

Response: The authors unified the vocabulary and changed the terms "clusters" to the term "classes". The sentence in the introduction has been modified. Furthermore, the LCA is now explained more detailed in the method section.

"The LCA is a probabilistic model which identifies mutually exclusive classes (risk groups) of a non-directly measurable variable. The LCA calculates the probability that a particular person belongs to a certain (latent) class as well as the probability that an individual in a certain class will give a particular answer [26]. These probabilities are known as unconditional class membership probability and class-specific response probability. The exact number of classes of the latent variable is not specified by the LCA." (p.7, l.7-12)

3. Please provide citations for your decision making around the number of classes you selected. The sentence "A low BIC, high entropy..." although non-controversial this should be back-up

by some literature. This is mainly to help readers understand why your decisions were made, and to guide researchers towards best practice.

Response: A reference has been added.

27. Geiser C (2010) *Data analysis with Mplus* [Original title: *Datenanalyse mit Mplus*]. VS Verlag für Sozialwissenschaften, Wiesbaden

4. Please define what you mean by “complex mixture term” as isn’t clear what it is or why you used from the reader’s perspective (end of page 6).

Response: The section has been revised to clarify the usage of the complex mixture term in MPlus.

“Since the sample of the ESA 2015 was drawn by a two-stage selection procedure, individuals were only selected at random within each sample point. The complex sampling of the ESA data required using a complex mixture term, since the results would be biased otherwise. Furthermore, all data was weighted (gender, age, region, and education) in order to provide results which are representative for the general adult population in Germany. The LCA was performed in MPlus using maximum likelihood estimation with robust standard errors [28]” (p.7, l.19-24)

5. From reading the next section “The conditional class membership probabilities were further used”, I believe you assigned each case to the class to which they had the highest probability of profile membership and then ran multinomial regression analysis. I would make this clearer as readers who are not knowledgeable about LCA may not understand this part. If you didn’t do this then please explain further.

Response: A general definition of the latent class analysis has been added for clarification.

“The LCA is a probabilistic model which identifies mutually exclusive classes (risk groups) of a non-directly measurable variable. The LCA calculates the probability that a particular person belongs to a certain (latent) class as well as the probability that an individual in a certain class will give a particular answer [26]. These probabilities are known as unconditional class membership probability and class-specific response probability. The exact number of classes of the latent variable is not specified by the LCA.” (p.7, l.7-12)

6. Please explain to the reader why you included a “complex mixture term to adjust standard errors” (page 7, first paragraph). It won’t be obvious to all readers

Response: The section was revised.

“Risk ratios as well as 95%-confidence intervals were calculated while using sampling weights and the svyset command to adjust for the complex sample design.” (p.7, l.31-33)

7. Please clarify in the same paragraph whether your multinomial regression analysis included all of your predictor variables (one model) or whether you entered each predictor univariately. From the notes in the table it is implied that you included together but please specify in this section of the method. Having split tables makes them easier to read, but also confuses the situation.

Response: Yes, all predictors were included at once and the authors added that information in the paragraph to avoid confusion. The tables were split to ensure readability. To avoid confusion, a note has been added below the tables stating “The results from the regression model are

presented in two tables (table2 and table 3) for reasons of clarity and comprehensibility, but only one regression model was calculated.” (p.11-12)

8. Class 3 – if higher scores on the ‘physical health’ item indicate poorer health, then are this class not more likely to report poor health?

Response: Thank you for your suggestion! The authors made a mistake and corrected it (p.9, l.24).

Reviewer: 5

Reviewer Name: Alan Watkins

Institution and Country: Swansea University, UK Competing Interests: None declared.

Summary: from the (necessarily) limited information provided in the manuscript, it would seem that the statistical analysis of the data in this study is appropriate, and that it has been competently performed - however, much more clarity is required on almost all statistical aspects of the manuscript, starting from a definition of latent class analysis (LCA) or, at least, a working description of its purpose – to identify an unknown (a priori) number of groups in a sample. The matter of what an LCA adds to one’s understanding also needs to be addressed more thoroughly.

Specific points requiring clarification in the manuscript and/or comment from the authors:

1. The role of weighting in the ESA sample needs to be clarified; the stated numbers and percentages in the “Sampling & procedure” paragraph need checking.

Response: The sampling weights were described in more detail in the data analysis paragraph. The stated numbers and percentages in the “Sampling & procedure” paragraph have been checked and were correct.

“Furthermore, all data was weighted (gender, age, region, and education) in order to provide results which are representative for the general adult population in Germany.” (p.7, l.22-23)

2. Given that LCA depends heavily (entirely?) on the risk factors, there needs to be better justification of those used here – if three (of eight) factors relate to cannabis, illicit drugs, and non-prescribed analgesics, it is hardly surprising that one latent class (albeit not identified by every modelling exercise) may be labelled “polysubstance use”.

Response: In order to address the issue and in accordance with reviewer 3 we have changed the label of class 4 to „Cumulate risk factors lifestyle“

3. Some of the definitions used to dichotomize risk factors seem arbitrary (as the authors admit); some result in thresholds that seem severe compared with others; and at least one needs checking – if I understand correctly, it is sufficient to undertake physical activity for less than 30 minutes per day on three days a week to be regarded as at risk as physically inactive, which is subtly different from what is written on page 6. This definition seems to set a very high threshold.

Response: The definition of physical activity has been revised to clarify the structure of the variable. Unfortunately, the ESA 2015 did not assess vigorous-intensity activity. Therefore, a statement has been added in the discussion, that the amount of physical activity might have been underestimated in the sample. As a consequence, the questions for assessing physical activity have been completely revised for the currently ongoing ESA 2018.

“For assessing physical activity, participants were asked on how many days per week they had been breathless and sweated because of physical activity within the last three months.

Participants, who were physically active for at least one day per week, were asked about the average duration of physical activity. The question could be answered on an ordinal scale (1 “Less than 10 minutes”, 2 “10 [PubMed](#) to less than 30 minutes”, 3 “30 [PubMed](#) to 60 minutes”, 4 “[PubMed](#) More than 60 minutes”). Based on the questions about frequency and duration of physical activity, a dichotomous variable was generated approximately fulfilling the recommendations of the American College of Sports Medicine and the American Heart Association [23]. Thus, physical activity for less than 30 min per day on five days per week was classified as a lifestyle risk factor.” (p.6, l.6-15)

“In addition, the ESA 2015 was only able to assess moderate-intensive physical activity and not vigorous-intensive physical activity, which could have led to an underestimation of physical activity in the sample. As a consequence, the questions for assessing physical activity have been completely revised for future assessments in ESA.” (p. 14, l.7-10)

4. Some sociodemographic factors seem to be very traditional – I was surprised that status did not explicitly account for “cohabiting” (conflate with “married”?) or “separated” (conflate with “divorce”?), as would almost certainly be done in the UK!

Response: The operationalization of the sociodemographic factors is conforming to the sociodemographic standards in Germany.

5. The paragraph on “Data analysis” on page 6 needs thorough revision, starting with a reference (such as Hagenaars & McCutcheon, 2002) for LCA. The method of estimation (EM algorithm?) needs to be mentioned, while discussion of starting values, iterations, convergence can either be omitted entirely, quickly summarised but without any technical terms (probably best for the BMJ), or further details should be provided – at the moment, the account falls between these stools. More details on the “complex mixture term” are definitely required!

Response: We added additional information on latent class analysis in general as well as on the method of estimation (maximum likelihood estimation with robust standard errors). Also, a reference has been added. Information about starting values, iterations and convergence have been omitted, as recommended by the reviewer. Furthermore, additional information on the complex mixture term has been added.

“A latent class analysis (LCA) was performed on eight categorical lifestyle risk factors as described above. The LCA is a probabilistic model which identifies mutually exclusive classes (risk groups) of a non-directly measurable variable. The LCA calculates the probability that a particular person belongs to a certain (latent) class as well as the probability that an individual in a certain class will give a particular answer [26]. These probabilities are known as unconditional class membership probability and class-specific response probability. The exact number of classes of the latent variable is not specified by the LCA. Goodness-of-fit measures (Bayesian Information Criterion (BIC), Entropy, Vuong-Lo-Mendell-Rubin adjusted Likelihood-Ratio-test (VLMR)) were used to select the best model. A low BIC, high Entropy and a significant VLMR indicate the best class-solution, but the final decision was also based on the researchers’ assessment of the interpretability of the results [27]. We tested models with up to five classes. Since the sample of the ESA 2015 was drawn by a two-stage selection procedure, individuals were only selected at random within each sample point. The complex sampling of the ESA data required using a complex mixture term, since the results would be biased otherwise. Furthermore, all data was weighted (gender, age, region, and education) in order to provide results which are representative for the general adult population in Germany. The LCA was performed in MPlus using maximum likelihood estimation with robust standard errors [28].” (p.7, l.6-24)

References:

26. Gollwitzer M (2012) *Latent-Class-Analysis*. In: Moosbrugger H, Kelava A (eds) *Test theory and questionnaire construction* [Original title: *Testtheorie und Fragebogenkonstruktion*]. Springer-Verlag, Berlin/ Heidelberg
27. Geiser C (2010) *Data analysis with Mplus* [Original title: *Datenanalyse mit Mplus*]. VS Verlag für Sozialwissenschaften, Wiesbaden

6. The second paragraph needs a reference to cover the multinomial logistic regression modelling; the reference to STATA needs to appear in the bibliography (cf the previous paragraph and Mplus).

Response: The reference to STATA now appears in the bibliography and a reference which covers the multinomial logistic regression has been added.

29. Berry WD, Feldman S (1985) *Multiple regression in practice*. Sage university paper series on quantitative applications in the social sciences. Sage, Beverly Hills, CA

7. Discussion of Table 1 needs to cover the use of weighting factors, and to compare weighted and unweighted percentages. This should be linked with the enhanced description of the ESA sample, previously mentioned.

Response: The authors chose not to show unweighted percentages, since only the weighted percentages are representative for the general adult population in Germany.

8. The reference to “bivariate residual statistics” needs clarification (may one assume that these are “standardised residuals”?), or, perhaps better here, a more reader-friendly phrasing, such as “Residual-based goodness-of-fit showed only minor violations ...”

Response: The sentence has been rephrased.

“Residual-based goodness-of-fit (standardized z-scores) showed only minor violations (> 1.96) of the assumption of local independence in all classes (data not shown) and therefore did not affect the selection of the final model.” (p.8, l. 6-9)

9. Table 2 could omit the first model (with only one class), and show subsequent changes in BIC relative to that model.

Response: Thank you for the suggestion. The authors decided not to change the table, since goodness-of-fit measures for LCAs are generally presented in this format.

10. More discussion need on the emerging classes, and the resulting partition of the ESA sample – of particular interest are those cannabis/illicit drug/non-prescribed analgesic users NOT categorised in class 4.

Response: Since the purpose of the LCA is to find classes with high probabilities of similar characteristics, the authors decided not to focus on single cases, which might not be categorized in class 4, because it would be beyond the scope of the paper. Furthermore, it is now mentioned in the statistics section, that the LCA only calculates the probabilities that a particular person belongs to a certain class. Additionally, the estimated class-specific response probabilities can be seen as high. If we did not answer the comment sufficiently we kindly ask for further guidance.

Finally, the standard of written English is generally good; in Tables 3 & 4 there is “und” rather than “and”.

Response: The sentence has been modified. (p.11-12)

VERSION 2 – REVIEW

REVIEWER	Sabine Rohrmann University of Zurich
REVIEW RETURNED	19-Jul-2018

GENERAL COMMENTS	All comments have been addressed appropriately.
-------------------------	---

REVIEWER	Rob Saunders UCL, UK
REVIEW RETURNED	05-Jul-2018

GENERAL COMMENTS	The authors should be commended for addressing all previous comments/queries. The manuscript is well written and describes a relatively complex analysis clearly for readers to be able to replicate and/or apply these methods in other datasets.
-------------------------	--

REVIEWER	Alan Watkins Swansea University, United Kingdom
REVIEW RETURNED	26-Jul-2018

GENERAL COMMENTS	This revised manuscript is a clear improvement on the previous version, but it is disappointing to note that some suggestions for changes to the previous version seem to have been ignored. I again urge the authors to provide more details behind the weightings applied to the sample points, or, if these are no longer readily available, to say so - it will, in my opinion, improve greatly the worth of the paper if these details can be made available. Tables 1 and 2 require particular attention; in the former, the raw and weighted percentages should be given, together with summaries of the weights involved for each risk factor. In Table 2, there are (presumably) both raw and weighted percentages for each class - these should be given. Then, there are corresponding changes to the accompanying text. Syntactical details (the svyest command) can be removed from the text; if though useful, the scripts used to run analyses can be made available, either from the corresponding author or as supplementary material.
-------------------------	--

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: Sabine Rohrmann

Institution and Country: University of Zurich, Switzerland Please state any competing interests or state 'None declared': None declared.

All comments have been addressed appropriately.

Response: Thank you very much.

Reviewer: 4

Reviewer Name: Rob Saunders

Institution and Country: UCL, UK

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

The authors should be commended for addressing all previous comments/queries. The manuscript is well written and describes a relatively complex analysis clearly for readers to be able to replicate and/or apply these methods in other datasets.

Response: Thank you very much.

Reviewer: 5

Reviewer Name: Alan Watkins

Institution and Country: Swansea University, United Kingdom Please state any competing interests or state 'None declared': None declared.

This revised manuscript is a clear improvement on the previous version, but it is disappointing to note that some suggestions for changes to the previous version seem to have been ignored. I again urge the authors to provide more details behind the weightings applied to the sample points, or, if these are no longer readily available, to say so - it will, in my opinion, improve greatly the worth of the paper if these details can be made available.

Tables 1 and 2 require particular attention; in the former, the raw and weighted percentages should be given, together with summaries of the weights involved for each risk factor. In Table 2, there are (presumably) both raw and weighted percentages for each class - these should be given. Then, there are corresponding changes to the accompanying text.

Syntactical details (the svyest command) can be removed from the text; if though useful, the scripts used to run analyses can be made available, either from the corresponding author or as supplementary material.

Response: Thank you for your review! The authors have provided more details behind the weightings applied to the sample points.

"Data was weighted using a redressement weight (age, gender, education, federal states and district size classes) based on the Iterative Proportional Fitting Algorithm [22]. The required marginal distributions of the population (18 to 64 years) were taken from the 2014 micro-census. The value of the redressement weight ranges between 0.19 and 5.63 (effectiveness 65.5% for $n = 5,962$) [21]." (p.5, l. 14-18)

Furthermore, an additional reference has been added: Gelman, A. and J. Carlin, Poststratification and weighting adjustments, in Survey nonresponse, R. M. Groves, J. L. Eltinge, and R.J.A. Little, Editors. 2002, Wiley: New York. p. 289–203.

Syntactical details (svyset command) have been removed from the text (p.7 l.33-34). The authors decided to make the script not available as supplementary material, since the authors conducted standard analyses, which are already described in the manual of each statistical software (Muthén & Muthén, 2010).

Table 1 has been completed with raw and weighted percentages for the lifestyle risk factors (p.8). All data has been weighted by gender, age, region (federal states and district size classes) and education, which is also mentioned below the table (p.8, l.12). If we have not answered the comment by the reviewer sufficiently, we kindly ask for further guidance.

Table 1

Raw and weighted prevalence rates for lifestyle risk factors ($n = 9204$)

Risk factors	Raw percentages		Weighted percentages	
	<i>n</i>	%	%	95%-CI
Daily Smoking	1322	14.5	18.5	[17.3,19.8]
At-risk drinking	1416	15.9	15.3	[14.4,16.3]
Episodic heavy drinking	2513	27.5	25.3	[24.0,26.7]
Cannabis use	386	4.2	3.1	[2.7,3.6]
Other illicit drugs use	75	0.8	0.8	[0.6,1.1]
Weekly pharmaceuticals use	733	8.0	8.2	[7.5,8.9]
Unhealthy diet	6185	67.4	69.9	[68.8,71.0]
Low physical activity	6423	84.1	82.6	[81.5,83.7]

Note. *n* = observed frequency of lifestyle risk factors, % = prevalence rates, 95%-CI = confidence intervals; weighted percentages: weighted for age, region, gender and education.

Thank you for your suggestion to include raw and weighted percentages for the class solutions in table 2. Since the data was drawn by a two-stage selection procedure it has been recommend by Muthén & Muthén (2010) to use a complex mixture term as method for estimation. Therefore, raw percentages for the class solutions have not been provided, because standard errors would be biased. The weighted percentages for the class solutions have been provided in the text (p.9, l.12, 15, 22 & p.10, l.3). The authors did not include the weighted percentages in table 2, since it would not improve the provided information on the class solutions and was not used for deciding the number of class. Furthermore, the presentation of the goodness-of-fit measures was based on Geiser (2010). If we have not answered the comment by the reviewer sufficiently, we kindly ask for further guidance.

Muthén, L. and B. Muthén, *Mplus User's Guide*. 6th ed. 2010, Los Angeles, CA: Muthén & Muthén.

Geiser, C., *Data analysis with Mplus [Original title: Datenanalyse mit Mplus]*. 2010, Wiesbaden: VS Verlag für Sozialwissenschaften.

VERSION 3 – REVIEW

REVIEWER	Alan Watkins Swansea University, UK
REVIEW RETURNED	17-Sep-2018
GENERAL COMMENTS	I think that the authors have now addressed my previous concerns as fully as possible, and I am content to recommend this manuscript for publication.