PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Smoking and suicidal behaviors in a sample of US adults with low mood: a retrospective analysis of longitudinal data
AUTHORS	Lirio S. Covey, Ivan Berlin, Mei-Chen Hu and Jahn K. Hakes

VERSION 1 - REVIEW

REVIEWER	John Hughes Professor, Univ of VT, USA
REVIEW RETURNED	14/02/2012

GENERAL COMMENTS	Review of “smoking and suicidal . . “ by Covey et al for BMJ Open Although many studies have associated smoking with suicide, whether this association is due to causal relations is unclear. The current study is one of the best tests of what this association means. It has many important assets including a very large sample, prospective design, representative sample, examination of both causal pathways and examination of changes in smoking and suicide outcomes, The results will likely have a significant effect on understanding of this association. The major drawback of the study is the large amount of jargon and the often unclear writing. I suggest the authors rephrase the objectives to be more clear and to avoid jargon terms such as “mutual causality” and “bidirectional relationships) and state they assess 1) whether smoking prospective predicts suicide, 2) whether suicidial behaviors prospectively predict smoking, 3) whether changes in smoking predict changes in suicide and 4) whether these relationships are due to psychiatric comorbidity. The authors are often confusing talking about “bidirectional” and “mutual causality” effects and seem to suggest one direction (smoking to suicide) is of more import than another. Why not just do two analysis (one for each direction of causality) and show both the same way – rather than report lots of data for one direction and little data for the other? In terms of jargon, please define “analytical, at risk sample,” “complementary regression” at first use. Technically former smokers are nonsmokers, so to be more clear I suggest substituting neversmokers for nonsmokers. I think the authors muddle confounds, mediators, and moderators. Much of the paper treats depression as confounder but then in discussion cites it as a mediator. To me, psychiatric disorders are very plausible mediators (remember psychiatrists believe mental disorders have a role in >90% of suicides). Thus analyses with them being considered as mediators rather than confounders is indicated at some point in the paper. I see no benefit to calling the sample smokers when in reality its tobacco users. Why not just call them tobacco users? I could not find the actual incidence rates for SROs (either unadjusted or adjusted) for the three groups. Would be best if actual

	fractions given. Also, what fraction was of each type (e.g. often ideation is 10-50 times more prevalent than attempts). Lines 20-29 on p 16 really confuse me. After reading several times, it seems to me it suggests suicide outcomes were taken among all Ss, not just those with low mood? When discussing duration of abstinence, I suggest you remind readers that probably very few of the recent quitters were in withdrawal and, thus, the current study is not an adequate test of whether tobacco withdrawal right after abstinence influences SRO I could not understand the logic of lines 27-41 on p 17 Abstract: As written objective 3 is unneeded as it is a consequence of the first two objectives. No need to cite Ss characteristics. Need to justify the use of low mood – e.g. state "who were asked about suicide" State outcome was over the last 3 years. I was surprised that after such state-of-the-art epidemiology that the authors made a causal conclusion. Prospective prediction is not causality Introduction. The authors should make more clear that the major issue in past papers has been whether the association is due to smoking being a marker vs being causal. The logic of the last sentence on p 5 and top of p 6 escapes me. I cannot see how using a subsample because you have to, increases validity. Methods We need a reference or website for NESARC. Again, need to clearly indicate why only those with low mood used. What were inclusion criteria? How was imputation done? I do not understand on p 10, how "survey design effects" influenced statistical significance. I suggest an exploratory analysis of each of the three outcomes by itself as the paper acknowledges outcomes vary by the dependent variables. Could be thought of as a sensitivity analysis We need the exact questions for smoking and suicide outcomes. Are diagnoses current or lifetime? I am unclear about the "tests for equality of coefficients" and no reference is given. Please explain What is the benefit of the Granger test over the simpler method of showing both directions analyses in a similar manner? The Granger test was published in 1969 and a literature search suggests it is rarely used. It is unclear if the test of SRO predicting smoking status uses smoking status at Wave 1 and other variables as covariates. Also was suicide asked only of those with low mood in Wave 1 as well. These are but two examples of how the analyses of this second causal path are not well described. Results: The interactions are hard to follow. I suggest first describing the interpretation (e.g. lines 29-39 on p 14) for each interaction before describing the results that lead to that interpretation P 15, line 15 – I cannot find this in Table 2 Discussion: The first sentence, needs to include the qualifier, "among those with low mood" Tables Are two decimal points really needed? Many of the "tables" are multiple tables and this is significantly

	confusing. I strongly suggest one table per labeled table. I would prefer to see the sample sizes in table 1 to help understand their validity. Table 2 is very enlightening. I do not understand why the (S) is entered. There is no way a reader could know the identity of the participant by just knowing the sample size. I strongly suggest the bottom part of Table 2 be a separate table with words instead of numbers in the first column. This is the most important table for the paper but is currently buried and hard to understand with numbers as descriptors. I am unclear about the "seventh" category – is this someone who starts as a never smoker or a former smoker? Table 3 could be put in text. The data in the second and third column is redundant with other tables. Table 4 is extremely hard to read. I suggest that the referent group (OR = 1.0) be put in upper left hand corner as most readers start their logic with the referent group and most read from left to right and top to bottom.

REVIEWER	John Stapleton Cancer Research UK Health Behaviour Research Centre, Department of Epidemiology and Public Health, University College London, UK
REVIEW RETURNED	15/03/2012

GENERAL COMMENTS	(1) The rationale for analysing a subset of Wave 2 responders is unclear. Whatever the reason, this cannot be described as a prospective study because outcomes are retrospective. A true prospective study would have looked at what happened to Wave 1 subjects wrt SRO's over the subsequent 3 years to Wave 2 follow-up. The only advantage of this design appears to be a 100% follow-up rate! (2) The analyses should be re-run without the original study weighting, which was presumable designed to estimate population prevalence. This is not the case here. In this re-analysis for a different purpose the adjusting background variables – age, gender, ethnicity etc. are the same variables as would have been used to weight the sample originally. Adjusting for these variables essentially re-weights the sample and it is unclear what the effect has been. (3) The results section proved quite tough to penetrate. One major problem was due to the frequencies of SRO's not being given for each of the characteristics in the tables. The heading to Table 1 implies that the first column does give frequencies, but this is not the case. Instead, they give the frequencies of the background characteristics. When readers look at the OR's they also need to see the frequencies they are based on to be assured that they make sense. These should be given in the form of percentages followed by the numbers i.e. 15.4% (n/N). This is standard good practice in medical journals. Other aspects of the layout and presentation could also be improved in-line with standard practice. (4) Table 3 is redundant. Part A is a repeat of previous results and the first 2 rows of Part B gives very obvious results. The only important result (AOR=0.81 (0.72-0.90) could be given in the text. Table 4 A is rather post-hoc and does not add significantly to the MS.

	(5) There are other interesting studies not cited e.g. Miller M. et al. Cigarette Smoking and Suicide: A Prospective Study of 300,000 Male Active dutyArmy Soldiers American Journal of Epidemiology 2000 151;11: 1060-3

VERSION 1 – AUTHOR RESPONSE

Reviewer: John Hughes
University of Vermont, USA, Psychiatry
Review of “smoking and suicidal . . “ by Covey et al for BMJ Open

Although many studies have associated smoking with suicide, whether this association is due to causal relations is unclear. The current study is one of the best tests of what this association means. It has many important assets including a very large sample, prospective design, representative sample, examination of both causal pathways and examination of changes in smoking and suicide outcomes, The results will likely have a significant effect on understanding of this association. The major drawback of the study is the large amount of jargon and the often unclear writing.
Response: We have meticulously edited the text and we hope that the revised version is devoid of jargon and unclear writing.
I suggest the authors rephrase the objectives to be more clear and to avoid jargon terms such as “mutual causality” and “bidirectional relationships) and state they assess 1) whether smoking prospective predicts suicide, 2) whether suicidial behaviors prospectively predict smoking, 3) whether changes in smoking predict changes in suicide and 4) whether these relationships are due to psychiatric comorbidity.
Response: We followed your suggestion and rewrote the objectives.
The authors are often confusing talking about “bidirectional” and “mutual causality” effects and seem to suggest one direction (smoking to suicide) is of more import than another. Why not just do two analysis (one for each direction of causality) and show both the same way – rather than report lots of data for one direction and little data for the other?

Response: We no longer use the terms “bidirectional” and “mutual causality”. We note that our data analysis placed more emphasis on the hypothesis that smoking causes SRO because it is more commonly asserted than the converse. We state in the Results that the models which tested effects of prior SRO on Wave 2 current smoking used the same control variables as the models on Wave 2 SRO. We also would like to draw attention to the fact that the latter relationship (SRO causes smoking) has not been tested in previous papers (with the exception of the Bronisch et al 2008 study that used an adolescent sample) and we thought that it should be.
In terms of jargon, please define “analytical, at risk sample,” “complementary regression” at first use. Technically former smokers are nonsmokers, so to be more clear I suggest substituting neversmokers for nonsmokers.
Response: We no longer use “analytical” and “complementary regression”. We now define the “at-risk sample” at first use as those reporting low mood at Wave 2 (page 8). As In all instances of its use, we substituted never smokers for nonsmokers.

I think the authors muddle confounds, mediators, and moderators. Much of the paper treats depression as confounder but then in discussion cites it as a mediator. To me, psychiatric disorders are very plausible mediators (remember psychiatrists believe mental disorders have a role in >90% of suicides). Thus analyses with them being considered as mediators rather than confounders is indicated at some point in the paper.
Response: Although, like the reviewer, we view depression and psychiatric disorders as likely mediators, we are unable to designate, parsimoniously, depression and other psychiatric disorders as

mediators as the timing of the psychiatric disorders relative to smoking within Waves is unclear. We take the conservative position of regarding depression and other psychiatric disorders (with possibly some exceptions such as ADHD), as confounders and have rewritten the manuscript to avoid muddling confounding, moderation, and mediation when referring to these predictors.
I see no benefit to calling the sample smokers when in reality its tobacco users. Why not just call them tobacco users?
Response: Only 129 persons of the sample of 7352 (page 8) reported using exclusively oral tobacco or snuff. Following NESARC procedures, we labeled these as smokers. A notable benefit would be data consistency with other NESARC papers reporting on smoking.
I could not find the actual incidence rates for SROs (either unadjusted or adjusted) for the three groups. Would be best if actual fractions given. Also, what fraction was of each type (e.g. often ideation is 10-50 times more prevalent than attempts).
Response: In Table 2, we show the weighted percentages of SROs reported in Wave 2 for each of the three Wave 1 smoking groups. We show in Table 1 weighted percentages of the individual SROs reported in Wave 1.
Lines 20-29 on p 16 really confuse me. After reading several times, it seems to me it suggests suicide outcomes were taken among all Ss, not just those with low mood?
Response: (Old) Lines 20-29 on p. 16 –

"Nevertheless, allowing for missing data from persons who did not report
depressed mood, further analysis on the total sample using the same regression model applied to the present, at-risk sample also showed significant, although lower, predictive effects of current and former smoking history on future SRO (AOR=1.36, 95% CI=1.25-1.48 and AOR=1.18, 95% CI=1.08-1.28, respectively).

We understand the reviewer's confusion. Those AORs using the total sample were based on models with the same control variables as the low mood sample, but that total sample regression was not valid as we had, in effect, added a lot of observations to the model that "necessarily" (following the skip-out of the SRO questions for persons with no low mood)) had "0"s for the dependent variable. We omitted that confusing sentence in the revised manuscript.

When discussing duration of abstinence, I suggest you remind readers that probably very few of the recent quitters were in withdrawal and, thus, the current study is not an adequate test of whether tobacco withdrawal right after abstinence influences SRO.
Response: This is true. We have added such a statement as a N.B. when describing former smokers in the Methods section, page 9.
I could not understand the logic of lines 27-41 on p 17.
Response: Lines 27-41 on p. 17 of the original manuscript are below:
"Former and nonsmokers with prior SRO in Wave 1 showed lower risk for current smoking in Wave 2 compared to their counterparts with no prior SRO. Perhaps those former and non-smokers, already inclined towards the more healthy behaviour of not smoking, were spurred by the prior SRO to undertake therapeutic actions which included avoidance of smoking. Of interest, exploratory analysis of Wave 1 data revealed a significant correlation (r=0.43, p=0.0001) between prior SRO and help-seeking during the last three years."

The logic underlying the statements involves the following: we have made the assumption that never smokers and former smokers are persons more inclined towards the healthier behavioral option of not smoking; we further conjectured that, the major experience of prior SRO spurred a portion of those subgroups to seek further therapeutic and health-promoting activities, and that avoiding future smoking is one of them. The serendipitous observation that prior SRO and treatment-seeking during the last three years were well correlated appears to be consistent with that conjecture. We have re-written the former lines 27-41 on page 17 to describe this thinking, now found on pp. 17-18.

Abstract:
As written objective 3 is unneeded as it is a consequence of the first two objectives. No need to cite Ss characteristics. Need to justify the use of low mood – e.g. state "who were asked about suicide" State outcome was over the last 3 years. I was surprised that after such state-of-the-art epidemiology that the authors made a causal conclusion. Prospective prediction is not causality
Response: We rewrote the Objectives; made them more concise. We removed the sentence on subject characteristics. Stating that "outcome was over the last 3 years" may be redundant because the three years of Wave 1 and 2 are stated. We agree, the word "consequence" is too strong. We rewrote the first sentence of the Abstract Conclusion to say "…smoking increased the risk of future suicidal behaviour…"

Nevertheless, we note in the final paragraph of the manuscript that the study findings of an independent, direct temporal effect of smoking on SRO and the negative temporal effect of SRO on smoking are more consistent with the hypothesis of a causal effect rather than a "bi-directional" relationship (sorry to use that term here) or the effect of a third shared factor.

Introduction.
The authors should make more clear that the major issue in past papers has been whether the association is due to smoking being a marker vs being causal. The logic of the last sentence on p 5 and top of p 6 escapes me. I cannot see how using a subsample because you have to, increases validity.
Response: The Introduction, first paragraph, now includes a sentence regarding the question of a correlational vs. causal relationship.
The implicit conjecture of the NESARC authors was that SROs are overwhelmingly experienced by those with low mood (in line with your previous remark "psychiatrists believe mental disorders have a role in >90% of suicides"). Our (necessary) use of the subsample increases statistical power and precision, discussed in the referenced article by Nordentoft M. 2007 Danish Medical Bulletin). In this sense, this "low mood" population is a sensitized population for SROs.
Methods
We need a reference or website for NESARC.
Response: We have provided the references to the Source and Accuracy statements for the NESARC Wave 1 and Wave 2 (reference #24, Grant et al, 2003, and reference #25, Grant et al, 2005, respectively), as well as the website references.
Again, need to clearly indicate why only those with low mood used. What were inclusion criteria?
Response: Please see our previous response as to the NESARC conception: the questions on SROs were asked because of the NESARC authors' assumption that SROs are largely related to low mood. SRO information (want to die, ideation, attempt) was not collected from persons who did not report low mood. We have also added the two low mood screening questions on page 6. Lifetime suicide attempt was asked of all respondents in Wave 2, but not want to die or suicidal ideation.
How was imputation done?
Response: We have added the following statement in the manuscript regarding the imputation method:
"Missing values, reported in Table 1, were replaced through imputation using assignment and allocation methods, as described in the Source and Accuracy Statements for Wave 1 of the NESARC (24,25)." (Page 11, bottom paragraph).

I do not understand on p 10, how "survey design effects" influenced statistical significance.

Response: Survey design effects (that is, stratification of the sampling frame and using clusters of observations from primary sampling units) reduce the variance of estimators relative to what one would get from simple random sampling. Smaller variances make tests of significance more sensitive,

and increase the power of the test to refute a false null hypothesis. This is also discussed in the Source and Accuracy Statements.

I suggest an exploratory analysis of each of the three outcomes by itself as the paper acknowledges outcomes vary by the dependent variables. Could be thought of as a sensitivity analysis

Response: In the Discussion, we acknowledge the remaining challenge of more complex analysis involving the individual SRO as a study limitation. We also mention findings from unadjusted analysis of similarities but also some differences in the relationships of smoking vis-à-vis the individual SROs and state that validation and articulation of these relationships require further analysis (p. 19).

We need the exact questions for smoking and suicide outcomes.

Response: We now provide the exact questions for smoking and SRO on page 9.

Are diagnoses current or lifetime?

Response: As stated in the original and in the revised versions, the psychiatric diagnoses are lifetime.

I am unclear about the “tests for equality of coefficients” and no reference is given. Please explain

Response: These are drawn from the Wald chi-square test, which is one of the first two tests of a joint hypothesis in a maximum likelihood model.

What is the benefit of the Granger test over the simpler method of showing both directions analyses in a similar manner? The Granger test was published in 1969 and a literature search suggests it is rarely used.

Response: The Granger procedure does test both directional analyses in a similar manner. For this paper, we used exactly the same set of control variables used to predict both Wave 2 SRO and Wave 2 current smoking: Wave 1 measures of demographics, socio-economic variables, psychiatric diagnoses, smoking status, and SRO.

The advantage of the test is in its simplicity. The right-hand side of the pair of models is identical, and the period 0 coefficient of X is used to test prediction of Y, and period 0 coefficient of Y is used to test prediction of X.

The Granger reference, which appeared in Econometrica, has been cited in 8053 other works. Please see (http://en.wikipedia.org/wiki/Granger_causality) for an on-line description.

It is unclear if the test of SRO predicting smoking status uses smoking status at Wave 1 and other variables as covariates. Also was suicide asked only of those with low mood in Wave 1 as well. These are but two examples of how the analyses of this second causal path are not well described.

Response: Yes, the test of SRO predicting Wave 2 smoking included smoking status (never, former, current) at Wave 1 and other covariates used in the model predicting Wave 2 SRO.

Yes, SRO was asked only of those with low mood in Wave 1 as well, stated in page 8. Respondents who did not report low mood in Wave 1 were assigned a value of 0 for prior SRO, stated in page 9.

Results:

The interactions are hard to follow. I suggest first describing the interpretation (e.g. lines 29-39 on p 14) for each interaction before describing the results that lead to that interpretation

Response: We have revised this section in the manner suggested by the reviewer. See page 15.

P 15, line 15 – I cannot find this in Table 2

Line 15 in former page 15: “smoking predicts increased risk of SRO but the reverse relationship does not hold”. In the original manuscript, the first part of the sentence is shown in the Table, but the reviewer is correct in that the latter part of the sentence is not shown. The revision corrects this by not referring the sentence to the Table.

Discussion:

The first sentence, needs to include the qualifier, “among those with low mood”

Response: The sentence is corrected in the revision.

Tables
Are two decimal points really needed?

Response: We noted that BMJ papers use 2 decimal points. Moreover, it’s a question of precision. The standard errors are accurate to the 1/100ths place and two significant figures. To limit them to one decimal point would often lead to sloppy construction of confidence intervals, such as when s.e. = 0.04.

Many of the “tables” are multiple tables and this is significantly confusing. I strongly suggest one table per labeled table.

Response: The sub-table below the old Table 2 has been eliminated; the data are described in the text. Old Table 4, which had two panels, has been revised to accommodate the data more meaningfully in one Table 4.

I would prefer to see the sample sizes in table 1 to help understand their validity.

Response: Table 1 (formerly, Supplemental Table 1) and Table 2 (formerly, Table 1), report the overall N for each table. The associated weighted percentages allow calculation of reasonably accurate estimates of each small n.

Table 2 is very enlightening. I do not understand why the (S) is entered. There is no way a reader could know the identity of the participant by just knowing the sample size.

Response: Old Table 2 is now Table 3, with “S” for Category 6 replaced by 95.

I strongly suggest the bottom part of Table 2 be a separate table with words instead of numbers in the first column. This is the most important table for the paper but is currently buried and hard to understand with numbers as descriptors.

Response: We no longer show the bottom sub-table in old Table 2. We now state the meaningful findings from that former sub-table in the text, with reports of the appropriate chi square and p values for each of the meaningful pair-wise comparisons. We think the new statements and the statistics shown in the results adequately the findings. (pp. 14-15).

I am unclear about the “seventh” category – is this someone who starts as a never smoker or a former smoker?

Response: The seventh category was comprised of persons who were never smokers in Wave 1 and self-reported as former smokers in Wave 2. Presumably, they began to smoke and stopped during the three years interval between Waves 1 and 2. Their numbers were too small for meaningful analysis. We state this on page 14 (last sentence, first paragraph) and in a caption below new Table 3.

Table 3 could be put in text. The data in the second and third column is redundant with other tables.

Response: Done. See last section of the Results section.

Table 4 is extremely hard to read. I suggest that the referent group (OR = 1.0) be put in upper left hand corner as most readers start their logic with the referent group and most read from left to right and top to bottom.

Response: We have re-arranged the data. The referent group of never smokers with no prior SRO (AOR=1.0) is shown in the first row of each sub-table.

Reviewer: John Stapleton
Cancer Research UK Health Behaviour Research Centre, Department of Epidemiology and Public Health, University College London, UK

(1) The rationale for analysing a subset of Wave 2 responders is unclear. Whatever the reason, this cannot be described as a prospective study because outcomes are retrospective. A true prospective study would have looked at what happened to Wave 1 subjects with SRO’s over the subsequent 3 years to Wave 2 follow-up. The only advantage of this design appears to be a 100% follow-up rate!

Response: We selected the subset of Wave 2 responders having decided that we wanted to predict future SRO rather than to conduct a follow-up study. Thus, Wave 2 outcomes could be predicted with Wave 1 variables.

It is true that SROs reported at Wave 2 are reported retrospectively. We agree that this is not a prospective study. We changed the title to describe the study as a “longitudinal analysis”.

(2) The analyses should be re-run without the original study weighting, which was presumable designed to estimate population prevalence. This is not the case here. In this re-analysis for a different purpose the adjusting background variables – age, gender, ethnicity etc. are the same variables as would have been used to weight the sample originally. Adjusting for these variables essentially re-weights the sample and it is unclear what the effect has been.

Response: With the weighting, we have a nationally-representative sample conditional upon having low mood, and the analysis produces estimators that are unbiased and consistent with respect to that conditional subpopulation. Without weighting, we have a sample that purposely over-represents certain subpopulations and geographic areas. This would have produced systematically-biased and inconsistent estimators if treated as a random sample (i.e. without weights). Please also refer to the referenced NESARC Source and Accuracy Statements.

(3) The results section proved quite tough to penetrate. One major problem was due to the frequencies of SRO’s not being given for each of the characteristics in the tables. The heading to

Table 1 implies that the first column does give frequencies, but this is not the case. Instead, they give the frequencies of the background characteristics. When readers look at the OR's they also need to see the frequencies they are based on to be assured that they make sense. These should be given in the form of percentages followed by the numbers i.e. 15.4% (n/N). This is standard good practice in medical journals. Other aspects of the layout and presentation could also be improved in-line with standard practice.

Response: The former supplemental data Table which describes the low mood sample and the rest of the sample (persons with no low mood) is now the new Table 1. This new Table 1 shows percentages by sub-categories of the covariates. Old Table 1 is now new Table 2 which shows weighted percentages of the SROs for each of the covariate sub-categories. As stated earlier, with our reporting of the overall N in Tables 1 and 2 and the conditional proportions of the covariates (Table 1) and Wave 2 SRO for each covariate (Table 2), a reasonably accurate estimate of each small-n is easily calculable by the reader (and reduces print space).

(4) Table 3 is redundant. Part A is a repeat of previous results and the first 2 rows of Part B gives very obvious results. The only important result (AOR=0.81 (0.72-0.90) could be given in the text. Table 4 A is rather post-hoc and does not add significantly to the MS.

Response: We suppressed old Table 3. The important result you mention is now in the text.

We re-formatted the data in Table 4, following the suggestion by John Hughes to start the data presentation with the reference group statistic AOR=1. We deliberated about excluding this table but finally chose to report it. We thought that the outcomes of the interactions between prior SRO and smoking status, while not fundamentally changing the observed relationships between smoking and SRO, revealed interesting, nuanced outcomes that could help to inform clinical decisions and further research.

(5) There are other interesting studies not cited e.g. Miller M. et al. Cigarette Smoking and Suicide: A Prospective Study of 300,000 Male Active duty Army Soldiers American Journal of Epidemiology 2000 151;11: 1060-3.

Response: We are aware of this study but did not include it among the references because its outcome is completed suicide whereas we focused on non-fatal suicidal behavior. Moreover, Miller et al (2000) did not control for psychiatric diagnoses which was a critical aspects of our analysis. In line with the character of our study, we cited only epidemiological studies of SRO, with the exception of Doll and Peto, 1976 because it is a classic as the first substantially-based study of smoking and suicide.

Thank you. We hope we have been adequately responsive to the reviewers' comments.

Lirio S. Covey, Ph.D.

VERSION 2 – REVIEW

REVIEWER	John Stapleton Cancer Research UK Health Behaviour Research Centre, Department of Epidemiology and Public Health, University College London, UK
REVIEW RETURNED	18/04/2012

THE STUDY	My major concern about the results concerns my previous point (2) regarding the statistical analysis. This has not been changed in the

	revision as suggested and the therefore the results may be unreliable. From their response I have the impression that the authors do not fully understand the statistical method they are using. To re-state: The data used have been previously weighted according to background characteristics – gender, age etc. for the purpose of obtaining unbiased population estimates of prevalence. This is not the purpose of the current analysis. When these same characteristics are used as adjusting variables, the new analysis effectively weights the sample again. It is unclear what the effect on the results of this double weighting has been. The analysis needs to be re-run without the original weights so that readers can be re-assured that the results are sound. Are the original un-weighted data available to the authors?
RESULTS & CONCLUSIONS	The authors have not addressed my previous point (1) adequately. Replacing “perspective” with “longitudinal” in the title does not help as this is simply semantics. The paper should be titled as a retrospective analysis from a longitudinal study. It must be a retrospective analysis study because the follow-up rate is 100%. Also, there is still not a clear explanation in ms as to why W1 respondents were not the basis of the analysis. The authors appear to have disregarding by previous point (3) concerning the frequencies of SRO’s being given for each of the characteristics considered eg males XX.X% (n/N), females YY.Y% (n/N). This is very poor practice and prevents the reader from having a clear perspective on the odds ratios that follow. Sensibly, leading medical journals have required this for many years and BMJ open should do likewise.

VERSION 2 – AUTHOR RESPONSE

Reviewer:

a) comment received April 12, 2012
My major concern about the results concerns my previous point (2) regarding the statistical analysis. This has not been changed in the revision as suggested and the therefore the results may be unreliable.
From their response I have the impression that the authors do not fully understand the statistical method they are using. To re-state:
The data used have been previously weighted according to background characteristics ? gender, age etc. for the purpose of obtaining unbiased population estimates of prevalence. This is not the purpose of the current analysis. When these same characteristics are used as adjusting variables, the new analysis effectively weights the sample again. It is unclear what the effect on the results of this double weighting has been. The analysis needs to be re-run without the original weights so that readers can be re-assured that the results are sound. Are the original un-weighted data available to the authors?

b) comment received on April 20, 2012
"Yes, all unweighted - analysis and reported figures (n/N). The weights in this survey, as all surveys of this type are included in order that estimates of population prevalence reflects the population. This analysis is quite different, and adjust for these same weight variables as best fits their modelling. The authors should re-run their analyses with unweighted data to check that the results are essentially the same (i.e. no changes in $p<0.05$). If they are, then they can simply make mention of this and report the analysis typical of other papers (likely correct, but do not have time to read them all)."

Response: We wish to point out that the weighting procedure performed on the total sample, and the

adjustment for covariates accomplished in the multiple regression are different tools, and are used for different purposes.

The NESARC weighting was for the purpose of correcting for oversampling and non-response and to allow generalizability to the US population as of the 2000 decennial Census. This weighting was to correct for sampling inequalities.

Thus, the study sample at Wave 2 served as a representative sample of the low-mood responders among the US population (Census 2000), made possible by the procedures that produced the NESARC weighted data used in the present analysis.

Performing the multiple regression on the unweighted data, would have achieved estimates that are not reflective of a representative sample of the persons with low mood (of the 2000 US population). The results would be skewed instead, encumbered by the oversampling (of young adults and Blacks and Hispanics) and nonresponse, and thus, questionable in its validity as a representative sample.

The inclusion of demographic variables in the regression models, on the other hand, serves to control for the covariance between demographic characteristics, the other regressors, and the dependent variable (SRO). Controlling for this covariation, which is the purpose of multiple regression, is necessary for unbiased estimation of the effects of the independent variables upon the dependent variable(s). This is true when the sample is proportional to all population demographics, and it is equally true when the sample has been reweighted to adjust for over-sampling in a survey design.

We note that all previous papers based on the NESARC used the same weighting method following NESARC guidelines. To not apply it for this study would be deviating from those guidelines and from the procedures used in other NESARC papers. Thus, in line with NESARC guidelines and other NESARC-based papers, we report weighted data in the manuscript.

In response to Dr. Stapleton's comment on April 20, 2012, we performed the multiple logistic regression models for assessing prediction of Wave 2 SRO and of Wave 2 current smoking based on unweighted data adjusted for design effects. We obtained similar results as the analyses based on weighted data (i.e., point estimates with statistical significance at $p<0.05$ and 95% confidence intervals not including 1).

Reviewer: The authors have not addressed my previous point (1) adequately.
Replacing ?perspective? with ?longitudinal? in the title does not help as this is simply semantics. The paper should be titled as a retrospective analysis from a longitudinal study. It must be a retrospective analysis study because the follow-up rate is 100%.

Response: This is true. We changed the title of the manuscript to indicate that the paper describes a retrospective analysis of longitudinal data. We made this change also in the Abstract.

Reviewer: Also, there is still not a clear explanation in ms as to why W1 respondents were not the basis of the analysis.

Response: Dr. Stapleton's request for a better explanation to the selection of the sample is well taken. We provide the requested explanation below.

We had considered several sample approaches, towards the aim of assessing prior smoking and covariates information (measured in Wave 1) as predictors of future (Wave 2) SRO. Thus, we wanted the sample that would enable us to utilize the most amount of reasonably reliable information, both in terms of number of respondents as well as in the availability of information from the most number of

respondents. A sentence stating this rationale has been added in page 8.

We found that the basic restriction to meeting those two goals was the NESARC design of asking about suicide related outcomes (SRO) only from persons with low mood. We decided that the current approach – using the sample that reported low mood at Wave 2, best met our desired analytic goals.

There were four potential samples – Sample 1 - persons with low mood at both Waves 1 and 2; Sample 2 - persons with low mood at Wave 1 followed in Wave 2, regardless of low mood status at Wave 2 (this is the prospective follow-up design); Sample 3 - persons with low mood at Wave 2 regardless of low mood status at Wave 1 (the present retrospective longitudinal analysis); and Sample 4 - the total sample, without restriction to low mood responders only.

a. The sample size for Sample 1 would have been 4,076.

b. The potential sample size for Sample 2 would have begun with a large number (11,340) of persons with low mood at Wave 1; however, since only persons with low mood were asked the SRO questions at both waves, the evaluable sample for predicting Wave 2 SRO would have been the same as for Design 1 (n=4076). There were 7254(2) persons from this initial low mood sample who reported no low mood in Wave 2 (thus, no SRO information). In addition to the loss of SRO information, there would also have been a lost utility of covariate information collected at Wave 1.

c. The sample size for Sample 3 , as pointed out by Dr. Stapleton – would have involved complete pick up of all reporting of SRO in Wave 2 from persons with low mood in Wave 2. Wave 2 low mood was reported by 7352 persons, regardless of low mood in Wave 1. This represents a pick-up (over Samples 1 and 2) of 3,276 persons (reported no low mood in Wave 1, but reported low mood in Wave 2). Importantly, these additional subjects from whom no SRO information was available in Wave 1 (and would have been excluded in a follow up study of only persons with low mood), were now evaluable as Wave 2 low mood responders, and Wave 1 covariate information for predicting the SRO outcome was also available.

d. Sample 4, i.e. the total Wave 2 sample and utilizing data from Waves 1 and 2, would have left too many SRO unknowns, or, if missing SRO was coded as 0, would have required too many assumptions that could not be justified. As part of this exercise for choosing the study sample, we used information collected for lifetime suicide attempt which was asked in Wave 2 but not in Wave 1, and found that there was a small percentage of lifetime suicide attempters (2.7%) who did not experience low mood either in Wave 1 or Wave 2.

Reviiewer: The authors appear to have disregarded my previous point (3) concerning the frequencies of SRO?s being given for each of the characteristics considered eg males XX.X% (n/N), females YY.Y% (n/N).
This is very poor practice and prevents the reader from having a clear perspective on the odds ratios that follow. Sensibly, leading medical journals have required this for many years and BMJ open should do likewise.

Response: n/N columns are now in each row of Tables 2-4. In Table 1, the percentages shown in each of the rows (within columns) are based on the same N at the top of each of the columns.

VERSION 3 – REVIEW

REVIEWER	John Stapleton Cancer Research UK Health Behaviour Research Centre, Department of Epidemiology and Public Health, University College London, UK

REVIEW RETURNED	09/05/2012

THE STUDY	Regarding my previous concern about the statistical analysis: (2) The analyses should be re-run without the original study weighting, which was presumable designed to estimate population prevalence. This is not the case here. In this re-analysis for a different purpose the adjusting background variables – age, gender, ethnicity etc. are the same variables as would have been used to weight the sample originally. Adjusting for these variables essentially re-weights the sample and it is unclear what the effect has been. The authors have refused to check the analysis using unweighted data and it is therefore unclear if the results are sound. This is a simple matter to check and I can only assume they have not done this because they do not have access to the raw data. This is unfortunate.

VERSION 3 – AUTHOR RESPONSE

In response to Dr. Stapleton's comment we did perform the multiple logistic regression models for assessing prediction of Wave 2 SRO and of Wave 2 current smoking based on unweighted data adjusted for design effects, and indicated that in the previous submission. We obtained similar results as the analyses based on weighted data (i.e., point estimates with statistical significance at $p<0.05$ and 95% confidence intervals not including 1). In the Methods, we report this finding as one of the sensitivity analyses chores, and indicate that analysis of the unweighted data, as the other sensitivity analyses conducted, did not alter the study findings.