PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (see an example) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

ARTICLE DETAILS

TITLE (PROVISIONAL)	A comparison of individual versus community influences on youth smoking behaviors: A cross-sectional observational study
AUTHORS	Adachi-Mejia, Anna ; Carlos, Heather; Berke, Ethan; Tanski, Susanne; Sargent, James

VERSION 1 - REVIEW

REVIEWER	M.R. Crone Associate Professor Leiden University Medical Center The Netherlands I have no competing interests
REVIEW RETURNED	28-Mar-2012

RESULTS & CONCLUSIONS	• An important addition to the discussion might be the findings of Pokorny et al. (Journal of Clinical Child & Adolescent Psychology, 2003) who found that a higher levels of exposure to tobacco retailers who sell to youth (measured by illegal retail tobacco sales to minors) was associated with smoking initiation. • At the end of the manuscript the authors mention that future studies should consider the role of density and proximity to tobacco outlets around schools. McCarthy (Am J Public Health 2009) looked at this aspect (adjusting for individual factors) and found a relationship between access to tobacco products in the proximity of the school and experimental smoking.
GENERAL COMMENTS	The aim of the present study was to assess the effects of community risk factors on adolescent smoking and to compare the added value to the effects of individual risk factors. The most important community factors were related to access to tobacco outlets: tobacco outlet density and tobacco outlet proximity. In conclusion the authors state that individual risk factors have a more important effect than these community factors. The most important factors for having tried smoking were sensation seeking and having friends smoking. Although it seems to be an interesting finding, the theoretical framework and the discussion of the results could be worked out more thoroughly. I have several additions: • The authors state that this study is the first to jointly examine individual risk factors with community-level factors: Pokorny and colleagues, however, looked at the combination of both groups of factors in an earlier study (Journal of Clinical Child & Adolescent Psychology 2003) • Even if the reported individual factors are important predictors of smoking behavior, the selection of included individual factors seems somewhat erratic and need some structuring.

	• It might be helpful to elaborate the description of the Theoretical Framework: how do the four systems of this framework influencing smoking behaviors relate to the factors measured in this study? The authors indicate that tobacco outlet is part of the meso system: however the model of Bronfenbrenner describes this system as the interrelations between micro systems. The authors describe the individual factors and, as far as I can see, factors from two systems: micro system (smoking siblings, smoking friends, perhaps also team sports participation), exo system (tobacco outlet, proportion of population black, proportion of low income etcetera). If using this framework, I also suggest using its categories/systems for reporting the results of the effects of the individual and contextual factors. • Considering that it is prohibited to sell tobacco products under the age of 18, one hypothesis might also be that there is a relationship between access to tobacco products and smoking intensity from the age of 18. Minor • Smoking intensity: how was it computed?
REVIEWER	Martine Shareck Ph.D Candidate Université de Montréal Canada I declare that I have no competing interests.
REVIEW RETURNED	22-May-2012

THE STUDY	No information is given regarding the representativity of the sample to the population of American youth. No information is given on missing data. Regarding variable description, a few clarifications could be made: p.6, line 13 – Please specify that adolescent's residential location was geocoded. Specifications could also be made elsewhere in the manuscript (eg. p.9). p.6, Smoking intensity : How was an overall measure of smoking intensity arrived at by combining 30-day smoking frequency and intensity on weekdays and weekend days ? p.7, Sociodemographics : More detail should be given on the creation of the socioeconomic index. It seems, from table 1, that this was a continuous variable, but this reviewer fails to see how parents' education and household income were combined into a continuous indicator of socioeconomic status. p.7, Sibling smoking : Response options should be given for this question since otherwise, readers will only understand later that this variable was used in a dichotomous (yes/no) expression. p.7, Team sports participation : This reviewer wonders about the relevance of using the number of sports team one has played on rather than the frequency (weekly or monthly for eg.) one has taken part in team sports. Also, was there a lot of variability in this variable ? If not, a dichotomous (yes/no) expression of team sports participation might be more meaningful and more relevant for smoking. Reading the results section, p. 14, it seems as if this

	variable was finally used as a categorical one. In any case, this should be clarified. p.8, Sensation seeking : How were responses for the 5 items combined into an overall index ?
RESULTS & CONCLUSIONS	Table 1 has a number of problems which make it difficult to understand : – Under the heading « tried smoking », proportion of participants having tried smoking, for a corresponding category, should be given rather than the « mean » – It is not clear what the « mean » for « smoking intensity » refers to – Which of these individual and community-level characteristics were statsitically significantly associated with either smoking outcomes ? These could be indicated with an asterisk and a footnote. – The interquartile range should be a signel number rather than a range, corresponding to Q3 – Q1. – Was friend smoking used as a continuous variable or a categorical one ? It appears in both sections of the table. p.18 : Please specify that proportion Black and Hispanics, and poverty were census-tract measures of these variables. Discussion section : – In the discussion, the authors start using the expressions « smoking onset » and « progression ». « Onset » and « progression » have a temporal connotation to it, which make them different concepts from what the authors have studied, i.e having ever tried smoking (with no indication of whether or not respondents went on to smoke regularly after trying) and intensity. This reviewer believes this distinction should be made clear and precise expressions should be used throughout the discussion. As well, further research could be suggested to address these issues, given the fact that the authors might have access to longitudinal data. - A critical question this reviewer has is the following : could the lack of associations between density and proximity of tobacco outlets in participants' residential area be due to the exposure variables and life environment (residential) studied not being relevant for this age group ? For example, sales to minors might be a more relevant exposure to investigate for young people who are not of legal age to buy tobacco products. As well, the school environment, or the combination of the residential and school environments might be more relevant spatial areas of exposure for this age group. These ideas are touched upon in the discussion, but this reviewer feels the conclusions (in the manuscript and in te abstract) should be nuanced in light of these hypotheses. – This reviewer is curious to know what type of environmental intervention the authors would recommend to lower sensation seeking among individuals (p.20, lines 15-17).
GENERAL COMMENTS	This article is well and clearly written. A major strength is the sophisticated methods used by the authors to estimate tobacco outlet density and proximity using kernel density estimates and road network distances respectively. Additional comments: A stronger rationale should be made for studying the variables

included in the models. The introduction/literature review is very short. Why would, for example, exposure to smoking in movies or participation in team sports be associated with higher or lower likelihood of having tried smoking or with smoking intensity?

Have the authors examined the interaction between age and neighbourhood-level exposures ? Could tobacco outlet density and proximity be more important for older age groups such as the 17 and 18 year-olds ? Investigating this interaction would be more in line with Bronfenbrenner's model which posits that different levels of risk factors, from the individual to the community, interact. This would be an interesting add up to the manuscript.

Specific comments :
p.4, line29 – This reviewer finds it misleading that the authors talk about « access » to tobacco outlets when actually referring to exposure to tobacco outlets in youth who, by definition, cannot legally purchase tobacco products. In this case, exposure does not necessarily translate into access and this distinction should be made clear

VERSION 1 – AUTHOR RESPONSE

Reviewer: M.R. Crone
Associate Professor
Leiden University Medical Center
The Netherlands
I have no competing interests

• An important addition to the discussion might be the findings of Pokorny et al. (Journal of Clinical Child & Adolescent Psychology, 2003) who found that a higher levels of exposure to tobacco retailers who sell to youth (measured by illegal retail tobacco sales to minors) was associated with smoking initiation.
>> We have added this reference to the discussion as suggested.

• At the end of the manuscript the authors mention that future studies should consider the role of density and proximity to tobacco outlets around schools. McCarthy (Am J Public Health 2009) looked at this aspect (adjusting for individual factors) and found a relationship between access to tobacco products in the proximity of the school and experimental smoking. The aim of the present study was to assess the effects of community risk factors on adolescent smoking and to compare the added value to the effects of individual risk factors. The most important community factors were related to access to tobacco outlets: tobacco outlet density and tobacco outlet proximity. In conclusion the authors state that individual risk factors have a more important effect than these community factors. The most important factors for having tried smoking were sensation seeking and having friends smoking. Although it seems to be an interesting finding, the theoretical framework and the discussion of the results could be worked out more thoroughly.
>> We have modified the manuscript in the Introduction, Methods, and Discussion to address the Reviewer's concerns. We have also added a reference to McCarthy et al. in the Discussion.

I have several additions:
• The authors state that this study is the first to jointly examine individual risk factors with community-level factors: Pokorny and colleagues, however, looked at the combination of both groups of factors in an earlier study (Journal of Clinical Child & Adolescent Psychology 2003)
>> We appreciate the Reviewer's suggestion to add this reference to the discussion in the point

above, and have added to our introduction that (unlike Pokorny et al.) our study is the first to engage in jointly examining these factors at the national level. Our study differs from Pokorny et al. in that we used geospatial analysis to measure tobacco retail density and proximity, and also differs from Pokorny et al. in that we control for many additional individual risk factors for youth smoking including exposure to movie smoking and sensation-seeking characteristics. Similar to Pokorny et al., we also did not find a relationship with tobacco retail density and youth smoking in our multivariate models. Their finding of a relationship between youth smoking initiation and youth purchases of tobacco in spite of the lack of relationship of youth smoking with tobacco retail density supports the need for a similar study to this one in young adults.

• Even if the reported individual factors are important predictors of smoking behavior, the selection of included individual factors seems somewhat erratic and need some structuring.
>> We have modified the introduction to clarify our selection of individual factors.

• It might be helpful to elaborate the description of the Theoretical Framework: how do the four systems of this framework influencing smoking behaviors relate to the factors measured in this study? The authors indicate that tobacco outlet is part of the meso system: however the model of Bronfenbrenner describes this system as the interrelations between micro systems. The authors describe the individual factors and, as far as I can see, factors from two systems: micro system (smoking siblings, smoking friends, perhaps also team sports participation), exo system (tobacco outlet, proportion of population black, proportion of low income etcetera). If using this framework, I also suggest using its categories/systems for reporting the results of the effects of the individual and contextual factors.
>> We have modified the methods to explain the framework more clearly in the context of our selected variables, and also added to the discussion our thoughts about the relevance of this framework to our findings.

• Considering that it is prohibited to sell tobacco products under the age of 18, one hypothesis might also be that there is a relationship between access to tobacco products and smoking intensity from the age of 18.
>> This is a good point and we agree that it should be made in the manuscript. We now emphasize in the limitations section that a null finding among adolescents does not necessarily imply no relation in populations that may legally purchase the product. We include a statement to that effect in the conclusion of the abstract as well. Understanding that a major target of the advertising is young adults, further study should include a similar study to this one among young adults only.

Minor

• Smoking intensity: how was it computed?
>> The score was created using the "alpha" command in Stata 12. This command computes the inter-item correlations or covariances for all pairs of variables in varlist and Cronbach's statistic for the scale formed from them. At least two variables must be specified with alpha. If used with the subcommand "generate(newvar)" the command specifies a scale constructed from the variable list be stored as a new variable. Items entered negatively are automatically reversed. If "std" is also specified, the scale is constructed by using standardized (mean 0, variance 1) values of the individual items. "Generate()" does not use casewise deletion. A score is created for every observation for which there is a response to at least one item (one variable in varlist is not missing). The summative score is divided by the number of items over which the sum is calculated. The variables included in the scale are described in detail in the methods section. In the case of this variable, we did not standardize the outcome measure.

Reviewer: Martine Shareck

Ph.D Candidate
Université de Montréal
Canada
I declare that I have no competing interests.

No information is given regarding the representativity of the sample to the population of American youth.
>> We did not mean to imply that this was a representative sample of American youth, only that it was a national sample. At baseline, the adolescents enrolled in the sample were similar to U.S. adolescents with respect to region of the country, age, gender (equal proportions), and family income. There were slight differences in terms of race (the study slightly under-represented Blacks and over-represented Hispanics). As mentioned in the methods section, by the 5th wave of the study (the basis for this manuscript) the sample was no longer representative, having lost higher proportions of minorities and the poor. For that reason, and in order to better study Black-White differences, the sample was "refreshed" with 598 Black adolescents. The sample, while not representative, captured adolescents from all 50 U.S. states.

No information is given on missing data.
>> There was little missing data. The number of adolescents surveyed was 3653, and the number in the final ever tried smoking model was 3543 (97%). We now state at the end of sample recruitment, "We used a complete case analysis approach because only 110 subjects (3%) were missing data from one or more variables in the analysis."

Regarding variable description, a few clarifications could be made:
p.6, line 13 – Please specify that adolescent's residential location was geocoded. Specifications could also be made elsewhere in the manuscript (eg. p.9).
>> On Page 6, we clarify that "Most adolescent residential locations were geocoded to their home street address (N=3167). When home street address was not available, they were geocoded to their ZIP code centroid (N=479)."

p.6, Smoking intensity : How was an overall measure of smoking intensity arrived at by combining 30-day smoking frequency and intensity on weekdays and weekend days ?
>> See response to Reviewer 1 above.

p.7, Sociodemographics : More detail should be given on the creation of the socioeconomic index. It seems, from table 1, that this was a continuous variable, but this reviewer fails to see how parents' education and household income were combined into a continuous indicator of socioeconomic status.
>> SES was created using the "alpha" command in Stata 12 (described above) and specifying that it should produce a standardized scale. If std is specified, the scale reliability is calculated as defined by the general form of the Spearman–Brown Prophecy Formula.(1, 2) This expression corresponds to holding the assumption that the summative rating is the sum of the standardized variables.(1) We now state, "The standardized variable is centered around zero and a 1-point increase corresponds roughly to a 1 standard deviation increase in the scale."

p.7, Sibling smoking : Response options should be given for this question since otherwise, readers will only understand later that this variable was used in a dichotomous (yes/no) expression.
>> We have made this addition.

p.7, Team sports participation : This reviewer wonders about the relevance of using the number of sports team one has played on rather than the frequency (weekly or monthly for eg.) one has taken part in team sports. Also, was there a lot of variability in this variable ? If not, a dichotomous (yes/no) expression of team sports participation might be

more meaningful and more relevant for smoking. Reading the results section, p. 14, it seems as if this variable was finally used as a categorical one. In any case, this should be clarified.
>> We regret not giving readers more information about this variable. We now include the following statement after the description of the ascertainment of the variable, "Team sports participation was skewed right, with responses ranging up to 12 but only four percent of responses greater than four. To limit outlier influence, all responses >4 were recoded to 4." The numbers correspond to the number of sports teams the subject reported playing on. As with the other variables, this one was recoded so that 0 corresponded to none and 1 corresponded to four teams, in order to allow an association size comparison with the other variables in the model.

We understand the Reviewer's point about frequency, but it was not asked about team sports.

p.8, Sensation seeking : How were responses for the 5 items combined into an overall index?
>> Sensation seeking was constructed using the "alpha" command in Stata 12 as described above. Since both reviewers inquired about this method, we added the following into the methods section just prior to the description of the dependent variables: "The dependent and independent variables include dichotomous, polychotomous and continuous variables. Three variables (smoking intensity, socioeconomic status, and sensation seeking) are scales derived from two or more items, constructed using the "alpha, gen(varlist)" command in Stata 12. The sections below describe how we ascertained the information, constructed the variables, handled outliers, and rescaled the variables in order to compare the associations in our analytical models."

Table 1 has a number of problems which make it difficult to understand :
– Under the heading « tried smoking », proportion of participants having tried smoking, for a corresponding category, should be given rather than the « mean »
>> We appreciate the Reviewer's point and have modified the text as suggested.

– It is not clear what the « mean » for « smoking intensity » refers to
>> Smoking intensity is a variable that ranges from 0 to 4. Mean refers to the mean level of the variable for the subjects in each group.

– Which of these individual and community-level characteristics were statsitically significantly associated with either smoking outcomes ? These could be indicated with an asterisk and a footnote.
>> We preferred not to include tests of statistical significance in Table 1, preferring to communicate this information instead in Tables 3 and 4, under the heading Unadjusted Odds Ratio. Presenting it in this way avoids arbitrariness of setting the p-value at 0.05 by presenting 95% confidence intervals.

– The interquartile range should be a signel number rather than a range, corresponding to Q3 – Q1.
>> We have modified the label in the table to clarify that we are showing Quartiles 1 and 3.

– Was friend smoking used as a continuous variable or a categorical one ? It appears in both sections of the table.
>> Friend smoking was used as a categorical variable; we have corrected this inconsistency in the table.

p.18 : Please specify that proportion Black and Hispanics, and poverty were census-tract measures of these variables.
>> We have added this specification to each of the tables.

Discussion section :
– In the discussion, the authors start using the expressions « smoking onset » and « progression ». « Onset » and «progression » have a temporal connotation to it, which make them different concepts

from what the authors have studied, i.e having ever tried smoking (with no indication of whether or not respondents went on to smoke regularly after trying) and intensity. This reviewer believes this distinction should be made clear and precise expressions should be used throughout the discussion. As well, further research could be suggested to address these issues, given the fact that the authors might have access to longitudinal data.

>> We appreciate the Reviewer's point and have modified the text to more accurately reflect what we studied, which is trying smoking and smoking intensity in a cross-sectional context.

- A critical question this reviewer has is the following : could the lack of associations between density and proximity of tobacco outlets in participants' residential area be due to the exposure variables and life environment (residential) studied not being relevant for this age group ? For example, sales to minors might be a more relevant exposure to investigate for young people who are not of legal age to buy tobacco products. As well, the school environment, or the combination of the residential and school environments might be more relevant spatial areas of exposure for this age group. These ideas are touched upon in the discussion, but this reviewer feels the conclusions (in the manuscript and
in te abstract) should be nuanced in light of these hypotheses.

>> We understand this point. Clearly, tobacco outlet density and proximity could have a null association with behavior in the context of an underage smoker, who is probably fairly constrained regarding where he or she may purchase tobacco. We clarify this point in the limitations section of the discussion. We added the following statement to the end of paragraph 3 in the discussion: "The lack of an association between tobacco outlet density and youth smoking does not preclude finding one in young adults or older adults, who are undoubtedly less constrained in their purchase of tobacco at retail outlets compared to youth." We added the following statement to the end of the abstract conclusion: "This finding may not apply to young adults, who are less constrained in where they may purchase tobacco."

We have added the role of the school environment as another area to consider as part of the limitations mentioned. The following statement may be found at the end of the fourth paragraph of the discussion: "Finally, we assessed tobacco density and proximity at the home, not the school. Future studies should consider the role of school density and proximity to tobacco outlets around schools in context of individual risk factors."

– This reviewer is curious to know what type of environmental intervention the authors would recommend to lower sensation seeking among individuals (p.20, lines 15-17).

>> We have published data that suggest seeing adult rated media increases growth in sensation seeking during adolescence. That study is cited in the discussion, and we repeat the citation here, along with the abstract of that publication (see Stoolmiller et al.).(3)

This article is well and clearly written. A major strength is the sophisticated methods used by the authors to estimate tobacco outlet density and proximity using kernel density estimates and road network distances respectively.

>> Thank you.

Additional comments:
A stronger rationale should be made for studying the variables included in the models. The introduction/literature review is very short. Why would, for example, exposure to smoking in movies or participation in team sports be associated with higher or lower likelihood of having tried smoking or with smoking intensity?

>> We have expanded the first paragraph to more clearly explain the variables included in the models.

Have the authors examined the interaction between age and neighbourhood-level exposures ? Could tobacco outlet density and proximity be more important for older age groups such as the 17 and 18 year-olds ? Investigating this interaction would be more in line with Bronfenbrenner's model which posits that different levels of risk factors, from the
individual to the community, interact. This would be an interesting add up to the manuscript.
>> Only 410 of the subjects were 18 or older, which limited the ability to test for an interaction. Nevertheless, we created an age dummy variable that was 1 for >=18 and 0 otherwise and entered it in each regression as an interaction variable with tobacco outlet density. We also created an age dummy variable that was 1 for >=17 and 0 otherwise and using the same method. Neither showed a significant interaction. We also reran both models after excluding subjects whose tobacco density was calculated at the Zip Code centroid. We have summarized these findings in the closing paragraph of the results section: "The tobacco outlet density—smoking association was not significantly different for those 18 years and older in either the tried smoking or the smoking intensity model. Additionally, excluding subjects for whom Zip Code centroid was used as a proxy for home address had little impact on the results."

Specific comments :
p.4, line29 – This reviewer finds it misleading that the authors talk about « access » to tobacco outlets when actually referring to exposure to tobacco outlets in youth who, by definition, cannot legally purchase tobacco products. In this case, exposure does not necessarily translate into access and this distinction should be made clear
>> We did not intend to mislead but were employing the more general meaning to the term "access" (Merrian Webster definition, "ability to enter, approach, or pass to and from a place or to approach or communicate with a person or thing"). Youth may not be able to legally purchase tobacco in retail outlets, but that does not mean they are barred from access to the multitude of messages emanating from these establishments, or that they cannot find a retailer willing to break the rules if they look hard enough. To clarify this concept, we have added an explanation about what we mean by access in the introduction. It reads, "This study examines the role of access to tobacco outlets compared to individual risk factors on youth smoking. By access, we mean approaching, entering, exiting, and having exposure to information imparted to potential customers about tobacco products, including visibility of in-store and storefront advertising"

REFERENCES

1. Nunnally JC, Bernstein IH. Psychometric Theory. 3rd ed. New York: McGraw-Hill; 1994.
2. Allen MJ, Yen WM. Introduction to Measurement Theory. Monterey, CA: Brooks/Cole; 1979.
3. Stoolmiller M, Gerrard M, Sargent JD, Worth KA, Gibbons FX. R-rated movie viewing, growth in sensation seeking and alcohol initiation: reciprocal and moderation effects. Prev Sci. 2010;11(1):1-13. Epub 2009/08/06.

STOOLMILLER, M., GERRARD, M., SARGENT, J. D., WORTH, K. A. & GIBBONS, F. X. (2010) R-rated movie viewing, growth in sensation seeking and alcohol initiation: reciprocal and moderation effects, Prev Sci, 11, 1-13.
ABSTRACT: The current study employed parallel process and discrete time hazard regressions to examine the interplay among exposure to R-rated movies, sensation seeking, and initiation of alcohol use in a national U.S. sample (N=6255) of adolescents, ages 10–14, who were followed over four

waves spanning 2 years. There was a short-term reciprocal
relation between watching R-rated movies and sensation seeking, but over the 2-year observation
period, exposure to R-rated movies was associated with increases in sensation
seeking and not vice versa. Sensation seeking also moderated the effect of watching R-rated movies
on initiation of alcohol consumption such that exposure was associated with greater increases in
initiation of alcohol use among low sensation than among high sensation seeking adolescents. The
study provides empirical evidence of an environmental media effect on sensation seeking, and
important new information about the relations among sensation seeking, media exposure, and
adolescent alcohol use.

VERSION 2 – REVIEW

REVIEWER	Shareck, Martine Université de Montréal, Médecine sociale et préventive
REVIEW RETURNED	20-Jul-2012

GENERAL COMMENTS	I was pleased to see that the authors clearly responded to the reviewers concerns and questions. The manuscript is now greatly improved. It is much clearer from the introduction why the specific risk factors included in the models were chosen. As well, I appreciate the definition given about what the authors mean by "access". I now only have a few minor comments and suggestions: - I understand that the scale used to represent smoking intensity combines days on which the participants smoked and the number of cigarettes smoked on those days. Higher values can thus either characterize the smoking intensity of an individual who has smoked a large number of cigarettes on a few days or few cigarettes on many days. I find this very interesting although the rationale for using this measure rather than more traditional ones such as average number of cigarettes smoked on days where someone smoked is not explicit in the manuscript. I think it would be an nice addition to include this and to show the advantages this measure offers. - there is a small mistake on p.7, line 46 : the sentence should be « those who responded « none » were categorized as never smokers » - the publication year of the Industry Classification documents used to estimate tobacco outlets measures should be given in the text to give a sense of how good a representation of youth's environments it offers. - What is the poverty level used to calculate the proportion of the population below the poverty level? This might be useful to compare this study with studies from other countries and it should be mentioned somewhere in the text. - there is a typo on p. 23, line 44: the sentence should read: "it would be unusual" rather than "an unusual" Martine Shareck PhD Candidate in Public Health and Health Promotion Université de Montréal

	Canada

VERSION 2 – AUTHOR RESPONSE

Reviewer: Martine Shareck
PhD Candidate in Public Health and Health Promotion
Université de Montréal
Canada

I was pleased to see that the authors clearly responded to the reviewers concerns and questions. The manuscript is now greatly improved. It is much clearer from the introduction why the specific risk factors included in the models were chosen. As well, I appreciate the definition given about what the authors mean by "access".

>> Thank you.

I now only have a few minor comments and suggestions:

- I understand that the scale used to represent smoking intensity combines days on which the participants smoked and the number of cigarettes smoked on those days. Higher values can thus either characterize the smoking intensity of an individual who has smoked a large number of cigarettes on a few days or few cigarettes on many days. I find this very interesting although the rationale for using this measure rather than more traditional ones such as average number of cigarettes smoked on days where someone smoked is not explicit in the manuscript. I think it would be an nice addition to include this and to show the advantages this measure offers.

>> We appreciate that the reviewer asked for this justification because it caused us to reconsider our approach. We have published using a similar measure in the past (Sargent JD, Hanewinkel R. Comparing the effects of entertainment media and tobacco marketing on youth smoking in Germany. Addiction. 2009 May;104(5)). However, in that manuscript, we used a combined variable created from lifetime smoking (How many cigarettes have you smoked in your life?' (none, a few puffs, one to 19 cigarettes, 20–100 cigarettes, more than 100 cigarettes)) and current smoking (How often do you smoke at present?' (I don't smoke, less than once a month, at least once a month but not weekly, at least once a week but not daily, every day)) to assess smoking intensity. The two items had good alpha reliability at 0.87. The three items in the present manuscript assess a slightly different outcome.

We think that predictive validity would be the best justification for any one formulation over another. We therefore went back to the data. This report is cross-sectional based on wave 5 data, but wave 6 smoking outcome data were available on about 900 of the almost 1300 ever smokers assessed at wave 5, wave 6 coming at approximately 1.5 years after wave 5.

For wave 6, we chose as a dependent variable a standard measure of 30 day smoking, which asked on how many days did they smoke in the past 30 and had the following distribution among the wave 5 eversmokers:
. tab cig30d6 if eversmk5==1
E14 / EA14: |
NUMBER OF |
DAYS SMOKED |
IN PAST 30 |
DAYS | Freq. Percent Cum.

```
--------------+--------------------------------
0: None | 519 57.99 57.99
1: 1-10 Days | 159 17.77 75.75
2: 11-29 Days | 50 5.59 81.34
3: Every Day | 167 18.66 100.00
--------------+--------------------------------
Total | 895 100.00
```

In an ordered logistic regression, we compared the ability of the three item scale used in this study (as mentioned in the methods section, the scale alpha reliability was 0.87) to predict wave 6 30-day smoking status against a two item scale similar to one used in the addiction paper cited above (scale reliability for the two-item scale among the ever smokers was 0.82). The results, shown below, reveal that the two-item predicts 30-day smoking 1.5 years later slightly better than the three-item measure. This was true even when other wave 5 covariates (HONC, smoker identity, smoking expectancies) were added to the predictive model (results not shown)

```
. ologit cig30d6 twoitem if eversmk5==1, or

Iteration 0: log likelihood = -982.15723
Iteration 1: log likelihood = -817.53081
Iteration 2: log likelihood = -812.09609
Iteration 3: log likelihood = -812.05416
Iteration 4: log likelihood = -812.05415

Ordered logistic regression Number of obs = 895
LR chi2(1) = 340.21
Prob > chi2 = 0.0000
Log likelihood = -812.05415 Pseudo R2 = 0.1732


-------------------------------------------------------------------------------
cig30d6 | Odds Ratio Std. Err. z P>|z| [95% Conf. Interval]
-------------+-----------------------------------------------------------------
twoitem | 4.687662 .4480021 16.17 0.000 3.886929 5.653352
-------------+-----------------------------------------------------------------
/cut1 | 2.014217 .1298302 1.759754 2.268679
/cut2 | 3.172667 .1594867 2.860079 3.485255
/cut3 | 3.671851 .1757176 3.327451 4.016251
-------------------------------------------------------------------------------

. ologit cig30d6 threeitem if eversmk5==1, or

Iteration 0: log likelihood = -982.15723
Iteration 1: log likelihood = -840.09356
Iteration 2: log likelihood = -836.94117
Iteration 3: log likelihood = -836.92325
Iteration 4: log likelihood = -836.92325

Ordered logistic regression Number of obs = 895
LR chi2(1) = 290.47
Prob > chi2 = 0.0000
```

Log likelihood = -836.92325 Pseudo R2 = 0.1479

```
--------------------------------------------------------------------------------
cig30d6 | Odds Ratio Std. Err. z P>|z| [95% Conf. Interval]
-------------+------------------------------------------------------------
threeitem | 5.021519 .5655172 14.33 0.000 4.026927 6.261759
-------------+------------------------------------------------------------
/cut1 | .8985167 .0834202 .7350161 1.062017
/cut2 | 1.994591 .1077589 1.783388 2.205795
/cut3 | 2.476372 .1232672 2.234772 2.717971
```

To keep things simple and consistent with existing literature, we have elected to substitute the two-item measure in the paper and have referenced the Addiction article as justification. We provide this additional justification in response to the substantive and insightful question raised by the reviewer.

- there is a small mistake on p.7, line 46 : the sentence should be « those who responded « none » were categorized as never smokers »
>> Thank you for pointing this out. We have corrected the text.

- the publication year of the Industry Classification documents used to estimate tobacco outlets measures should be given in the text to give a sense of how good a representation of youth's environments it offers.
>> The tobacco outlet data came from year 2007. We have added this to the text on page 10.

- What is the poverty level used to calculate the proportion of the population below the poverty level? This might be useful to compare this study with studies from other countries and it should be mentioned somewhere in the text.
>> We have found that comparing poverty indexes across countries is a complex issue. As you can see from the following website, the United States Census Bureau uses a complicated process to measure poverty: http://www.census.gov/hhes/www/poverty/methods/measure.html
There are 48 possible poverty thresholds that vary according to family size and ages of household members. Therefore, we find a comparison across countries to be beyond the scope of this manuscript.

- there is a typo on p. 23, line 44: the sentence should read: "it would be unusual" rather than "an unusual"
>> We have made this correction.

I now believe that the manuscript is acceptable for publication given the small corrections asked for in my review. Thank you.
>> Thank you.