

## 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

<b>TITLE (PROVISIONAL)</b>	National medical expenditures by smoking status in American adults: an application of Manning's two-stage model to nationally-representative data
<b>AUTHORS</b>	Swedler, David; Miller, Ted; Ali, Bina; Waehler, Geetha; Bernstein, Steven L.

### VERSION 1 - REVIEW

<b>REVIEWER</b>	Sapna Kaul University of Texas Medical Branch, USA
<b>REVIEW RETURNED</b>	22-Oct-2018

<b>GENERAL COMMENTS</b>	<p>Overall comments: As mentioned by the authors, this research topic has been explored previously and the current attempt was to provide an update on these findings perhaps using a different method. A recent paper (Xu et al., 2015, AJPM, <a href="https://www.sciencedirect.com/science/article/pii/S0749379714006163">https://www.sciencedirect.com/science/article/pii/S0749379714006163</a>) used the NHIS linked MEPS data to provide information on smoking attributable costs among U.S. adults, which is cited in this manuscript. They found that the annual costs amounted to \$170 billion per year (which is higher than the number projected in the current manuscript, \$101 billion). This paper therefore does not add any new insights or novel findings.</p> <p>Below, I comment specifically on areas that need attention.</p> <p>Abstract Conclusions: This is confusing – “Former smokers had lower medical expenditures than Current smokers in all age groups. While we identify Former smokers as having higher medical expenditures than Current smokers, we do not examine how care-seeking behavior varies between levels of each risk factor.” The former line suggests former smokers had lower expenditure the latter says they had higher expenditures – which is correct?</p> <p>Introduction Page 4 Line 12: please describe the “AF method” briefly for those who do not understand this method. Line 19: please provide citations for “health economics literature has shifted from this AF method ....”. For this statement, Manning et al.'s seminal paper on generalized modeling approaches is cited.</p> <p>Page 5</p>
-------------------------	---

	<p>Lines 1 to 4 – this section is confusing: how does this method account for coding errors (a well visit being coded as a chronic visit)? The standard errors capture the uncertainty around the sample means. If the sample size were to increase, the uncertainty will go down but the coding errors will probably remain. Can you please explain in more detail what you mean by the coding issue that you brought up?</p> <p>Lines 11 to 14: How about the impact of socio economics? Poorer individuals may be more likely to engage in risky behaviors and also be less likely to see doctors.</p> <p>Page 6 Lines 8 to 11: Please provide more details on what all expenditures were used. Did you use all expenditures (outpatient, inpatient, dental etc.) – out of pocket and those billed to insurance, and were these self-reported? Were the expenditures adjusted for inflation using any index?</p> <p>Page 8 Lines 2 to 3: Did you use OLS or logistic as the first step? In the second step, did you specify a link for GLM? Lines 5 to 7: How were the survey weights used in “twopm”? Did you use the “svy” set commands?</p> <p>Line 13: this sentence seems to be missing impact or effect. “(Including it had negligible on estimated expenditures by smoking status.)”</p> <p>Lines 19 to 20: Was this model only estimated for former smokers?</p> <p>Page 10 Lines 6 to 12: Are these numbers derived by extrapolating your average costs? If yes, what was the reason behind extrapolating the annual cost (multiplying the average cost by average number of smokers)? Is it not possible to use your data to calculate the annual/total expenditure for current, former and never smokers?</p> <p>Line 23: Why would one observe lower marginal or average estimates in MEPS? If NHEA includes institutionalized adults, unless the costs of health care for these adults are very different, the marginal/average should not be very different? I understand that when you extrapolate, the total costs could be different because of a different population base.</p> <p>Page 13 Lines 16 to 19: While a dummy variable for survey year may adjust for potential temporal effects, will the average/total cost estimates still reflect values in current dollars?</p>
--	--

<b>REVIEWER</b>	Kenshi Hayashida University of Occupational and Environmental Health, JAPAN
<b>REVIEW RETURNED</b>	29-Jan-2019

<b>GENERAL COMMENTS</b>	<p>This study aims to examine annual medical expenditures for American adults based on current smoking status.</p> <p>Major comments:</p>
-------------------------	---

	<p>1. Novelty or Originality The strength of this study is using the latest data and the method that seems to be the best now. However I cannot feel novelty or originality other than them.</p> <p>2. Number of data The numbers of data used in Tables 3 and 4 are unclear. Add figures of each category (1-, 2-, and 5-year thresholds) in Table 3, and that of each decade in Table 4. Moreover, if the number of people (proportion) in each group seems to be biased, discuss the influence on the results.</p> <p>Minor comment:</p> <p>1. Page 6, line 20- 22 It seems that the same variable (e.g. educational level and educational data) is described as covariates. Please check and correct the sentences as necessary.</p>
--	--

### VERSION 1 – AUTHOR RESPONSE

Reviewer #1

Abstract

This is confusing – “Former smokers had lower medical expenditures than Current smokers in all age groups. While we identify Former smokers as having higher medical expenditures than Current smokers, we do not examine how care-seeking behavior varies between levels of each risk factor.” The former line suggests former smokers had lower expenditure the latter says they had higher expenditures – which is correct?

We deleted the sentence “Former smokers had lower medical expenditures than Current smokers in all age groups.”

Introduction

Page 4

Line 12: please describe the “AF method” briefly for those who do not understand this method.

Line 19: please provide citations for “health economics literature has shifted from this AF method ....”. For this statement, Manning et al.’s seminal paper on generalized modeling approaches is cited.

We added citations to support the statement that the Manning method has been increasingly in favor. The Manning citation was then moved to the following sentence.

The AF method has been further described from Rockhill et al’s review article: “In their review of the AF method, Rockhill and colleagues define attributable fraction and similarly-phrased terms to be ‘the proportion of disease risk in a population that can be attributed to the causal effects of a risk factor or set of factors,’ (p. 15) [10]. Rockhill et al find that the AF method has limitations, including that is constrained by how accurately costs are allocated among diagnoses and by the availability of accurate AFs that are not confounded by co-occurring risk factors.”

Page 5

Lines 1 to 4 – this section is confusing: how does this method account for coding errors (a well visit being coded as a chronic visit)? The standard errors capture the uncertainty around the sample means. If the sample size were to increase, the uncertainty will go down but the coding errors will probably remain. Can you please explain in more detail what you mean by the coding issue that you brought up?

The coding of a specific visit was not our focus here, and we can see why the reviewer was confused by our language. We have elaborated on how the two-stage model associates medical expenditures broadly with risk factors: “That is, these equations take a more holistic view of medical care and expenditures associated with risk factors’ coefficients than would a process seeking to attribute specified treatment costs to specified factors. It allows that medical visits for conditions related to the behavioral risk factors may displace other medical care that might be sought if an individual did not have a given risk factor.”

Lines 11 to 14: How about the impact of socio economics? Poorer individuals may be more likely to engage in risky behaviors and also be less likely to see doctors.

The reviewer is correct that SES affects medical expenditures. There are in fact many other factors, such as medical history and race/ethnicity, that impact care seeking. Because our focus is so narrowly on costs associated with particular behaviors, we have kept this paragraph focused on how smoking may impact care-seeking. A broader study on the impact of all these factors on care-seeking would be a great, but highly involved, follow up to our work.

Page 6

Lines 8 to 11: Please provide more details on what all expenditures were used. Did you use all expenditures (outpatient, inpatient, dental etc.) – out of pocket and those billed to insurance, and were these self-reported? Were the expenditures adjusted for inflation using any index?

On lines 6-7 we state that all payments are tracked, regardless of payer/source. We added information on the PCE inflators used to bring all costs to 2015 US\$.

Page 8

Lines 2 to 3: Did you use OLS or logistic as the first step? In the second step, did you specify a link for GLM?

Lines 5 to 7: How were the survey weights used in “twopm”? Did you use the “svy” set commands?

Thank you for noticing that these were not included. OLS used probit and the second equation had a log link. Both details were added to the text.

We state later in the paragraph that the ‘svy’ command was employed with weights provided by AHRQ.

Line 13: this sentence seems to be missing impact or effect. “(Including it had negligible on estimated expenditures by smoking status.)”

Thank you for catching this. The parenthetical now reads “(Further sensitivity analysis indicated that including body mass had negligible impact on estimated expenditures by smoking status.)”

Lines 19 to 20: Was this model only estimated for former smokers?

We have clarified that this analysis on years-since-quitting was an expansion on Model 2. Years-since-quitting is only applicable to Former smokers.

Page 10

Lines 6 to 12: Are these numbers derived by extrapolating your average costs? If yes, what was the reason behind extrapolating the annual cost (multiplying the average cost by average number of smokers)? Is it not possible to use your data to calculate the annual/total expenditure for current, former and never smokers?

This is correct: we multiplied individual costs by the numbers of Current, Never, and Former smokers. Since our unit of analysis was the individual, we did not re-run our models to examine population-level smoking costs. We have added 95% CIs to our estimates to reinforce that these data have measurement uncertainty around them.

Line 23: Why would one observe lower marginal or average estimates in MEPS? If NHEA includes institutionalized adults, unless the costs of health care for these adults are very different, the marginal/average should not be very different? I understand that when you extrapolate, the total costs could be different because of a different population base.

We have expanded upon the list of groups covered by the NHEA but not MEPS: “However, the NHEA is more comprehensive than the MEPS in capturing Medicaid costs covering institutionalized adults (including those in nursing homes), active-duty military, and foreign visitors to the US 28 29, so we expected MEPS to yield lower estimates”. We can understand why our initial text didn’t convey how different the two populations covered by the databases were.

Page 13 Lines 16 to 19: While a dummy variable for survey year may adjust for potential temporal effects, will the average/total cost estimates still reflect values in current dollars?

The costs are in 2015 USD. We have clarified this in multiple places within the text.

Reviewer #2

The numbers of data used in Tables 3 and 4 are unclear. Add figures of each category (1-, 2-, and 5-year thresholds) in Table 3, and that of each decade in Table 4. Moreover, if the number of people (proportion) in each group seems to be biased, discuss the influence on the results.

Thanks for pointing out that we lacked this information. In the first Results paragraph, we added a sentence saying “For the average 23.9 million Former smokers, only 4.3% quit within the prior year, 8.7% quit within the prior 2 years, and 24.4% quit within the 5 years prior to the survey.” This is approximately 4.5% of smokers quitting each year.

Page 6, line 20- 22: It seems that the same variable (e.g. educational level and educational data) is described as covariates. Please check and correct the sentences as necessary.

Thanks for catching this duplication. We have corrected it.

#### **VERSION 2 – REVIEW**

<b>REVIEWER</b>	Kenshi Hayashida University of Occupational and Environmental Health, Japan
<b>REVIEW RETURNED</b>	02-Apr-2019

<b>GENERAL COMMENTS</b>	Thank you for the respond to the reviewer's comment appropriately. However I cannot feel novelty or originality as commented previously, because this paper has no new findings unfortunately.
-------------------------	---