PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Geographic variation in tobacco use in India: A population based multi-level cross-sectional study
AUTHORS	Singh, Ankur; Arora, Monika; Bentley, Rebecca; Spittal, Matthew; Do, Loc; Grills, Nathan; English, Dallas

VERSION 1 – REVIEW

REVIEWER	Fiore Maria Department "GF Ingrassia", University of Catania, Italy
REVIEW RETURNED	11-Aug-2019

GENERAL COMMENTS	The manuscript is well done, but figure 1 is not clear.

REVIEWER	Amit Yadav Center for Tobacco Control Research and Education (CTCRE), University of California, San Francisco, California, The United States of America
REVIEW RETURNED	21-Dec-2019

GENERAL COMMENTS	Congratulations on this great effort to present the geographical variability of tobacco use in India. This analysis from the GATS will surely help the policy makers at the national, state and grass roots to focus on targeted approach to implement evidence based and effective tobacco control measures in the country. While the overall structure and design of the paper is unique a relook to make it little jargon free with more layman explanations for some of the technical terms used in the paper e.g. ‘atomistic fallacy’, ‘proportionate universalism’ etc. will make it more reader friendly. Some of the specific comments on the paper that may be considered by the authors in revising the paper are as follows: **Method** The paper is based on Multilevel Modeling approach to account for the variability in different kind of tobacco use at city/village levels using the multistage sampling design of GATS. However, it is not clear as to how authors have decided that multilevel modeling is better than the logistics regression? Statistical tests (like LR) can be used to justify the purpose of using multilevel modeling over logistics regression. Not using the policy and economic variables puts a serious limitation on the study as the tobacco use in any area will be greatly affected

	by the level of tobacco control policy implementation and the economic capacity of population in that area. Even if in the author's estimation this may result in 'atomistic fallacy', it would be good to include it for the readers to get a better understanding of the actual scenario as the information is based on the same data source. While the main purpose of paper is to determine the spatial variation in tobacco use at ward and village level, it is not clear why states have not been included in the models which is likely to effect the analyses significantly. Page 2 line 18 and pg5 line 58: Total sample size of GATS-2 was 74,037 individuals from 30 states and two union territories and not 31 states and union territories. **Results** It is not clear how the results have been controlled for the impact of state, district and local level tobacco control policy and implementation of COTPA. It would be more helpful if the geographical variations are compared and presented for GATS-1 vs GATS-2. In addition if this is further looked from an NTCP v. non-NTCP districts/ward/village that will also help in arriving at more specific policy recommendations for strengthening NTCP at local level. It is suggested to provide full model details with all independent variables with corresponding OR, P-value and 95%CI for the benefit of readers. Table 1 (descriptive statistics) does not include urban-rural but mentioned in remaining table footnotes. Please clarify whether subsequent multilevel models are adjusted for urban-rural? As mentioned previously, it is much better to provide the full details of all tables rather than just writing in the footnote for better clarity, in case of space limitation it can be included as online supplement. Pg 10 line 6: Not able to follow the mentioned urban rural variance in the table-2 Pg 10 line 10: what was the median odds ratio for rural areas? **Discussion** It would be helpful for the readers if the discussion adds on what are the possible reasons for higher clustering of tobacco use at area level? And also the possible reasons for higher clustering of dual use and SLT use? Whether data from GATS-1 and GATS-2 are unable to explain the factors responsible for the high variations of tobacco use at individual level or area level or both? How this limitation can be overcome? Please give some examples of area-level determinants that is being suggested to be included for future GATS surveys.

	Policy implications of the paper can be further discussed keeping in mind the structure of governance at the national/state/district/local level and compliance with NTCP, COTPA and other tobacco control policies at these different levels. "Comparison of GATS-2 and GATS-1 has highlighted changes in prevalence of tobacco use due to differential implementation of these measures." (Reference?) "States are also allowed to develop context specific information, education and communication resources to match the local needs." (Refrence?) It will be helpful for the readers to have some idea of the current design and reach of the National Tobacco Control Programme at district, block and village level? **Table** Table 1 on page 18 has two categories of dual use which is confusing and non-confirming to the definition of dual use in the paper.

REVIEWER	Neal Freedman NCI, USA
REVIEW RETURNED	03-Feb-2020

GENERAL COMMENTS	The authors present an interesting analysis of the impact of geographic location on the prevalence of tobacco products in India. I think that the manuscript adds to the literature and provides important information. Yet, I feel that the manuscript would benefit from revision in several areas. 1) As currently written and analyzed, it is hard to understand how the underlying prevalence data is reflected in the statistical models. For this reason, I think the authors should provide the prevalence of each tobacco product in each geographic location included in the analysis for readers. This could be done via a table and a map. For example, how does the prevalence of each type of tobacco & dual use vary across the included geographic areas? 2) Then, I think it would be useful for the authors to further justify and explain the advantages of their approach versus simply examining the prevalence of each tobacco product in each geographic location using simpler methods. And further, how is the information gained from these analysis help make public health decisions. For example, one approach might be to target areas with high prevalence—or alternatively, areas with a large # of users. How would the information gained by the analyses described herein contribute to such decisions? Or would it? 3) The abstract should spell out all abbreviations, such as MOR, and further should detail the methods used. 4) Results, page 10, top paragraph: This conclusion seems overstated. As elsewhere, tobacco use typically begins in India in youth—thus it seems very unlikely to me that a middle-aged individual who doesn't use tobacco would begin because they moved to a new area. Perhaps geographic differences are most correlated in rates of initiation and cessation?

	5) The authors suggest that future GATS surveys should add questions in order to assess area-level determinants (page 12). I think this statement would benefit from clearly describing what sorts of questions should be added.

VERSION 1 – AUTHOR RESPONSE

Reviewer# 1

Comment 1: The manuscript is well done, but figure 1 is not clear.
Response: We have added the following text to the results section to make the figure clear:

> For each outcome, the AUC increased when multilevel logistic regression models were fitted. The AUC increased to 0.86 with Model B as compared to 0.79 in single-level logistic regression model (Model A) implying the presence of a general contextual effect and the ability to better classify individuals according to tobacco use (Figure 1). Changes in AUC were highest for smokeless tobacco use 11%, compared to 2% for cigarette smoking (Table 2 and 3).

Reviewer# 2

Comment 1: Congratulations on this great effort to present the geographical variability of tobacco use in India. This analysis from the GATS will surely help the policy makers at the national, state and grass roots to focus on targeted approach to implement evidence based and effective tobacco control measures in the country.
Response: Thank you.

Comment 2: While the overall structure and design of the paper is unique a relook to make it little jargon free with more layman explanations for some of the technical terms used in the paper e.g. 'atomistic fallacy', 'proportionate universalism' etc. will make it more reader friendly.
Response: We have thoroughly edited the paper to make it more accessible. We have now explained the terms within the text:

Atomistic fallacy:

> We did not incorporate policy and economic variables related to tobacco use available in the GATS 2016-17 in our analysis because the policy and economic variables were respondent's perceptions rather than objective measures of availability and implementation of policies in local areas. The non-ecologic nature of these variables could lead to falsely attributing individual-level measures to area levels (the atomistic fallacy).

Proportionate universalism (we have changed the text to make it more accessible):

> Finally, our use of the multilevel approach in this study advances a 'proportionate universalism' approach suggesting tobacco control interventions applied nationally should be scaled according to local area level disadvantage to reduce geographic inequalities.

Comment 3: (Method) The paper is based on Multilevel Modeling approach to account for the variability in different kind of tobacco use at city/village levels using the multistage sampling design of GATS. However, it is not clear as to how authors have decided that multilevel modeling is better than the logistics regression? Statistical tests (like LR) can be used to justify the purpose of using multilevel modeling over logistics regression.

Response: Thank you for your comment. Our decision to use multilevel modelling techniques were based on the following two considerations:

- ***i) Conceptually, multilevel models explicitly partition variation in an outcome in terms of individual characteristics and their social and political context. This addresses our paper's key objective of quantifying overall variability in tobacco use across city wards and villages.***
- ***ii) Statistically, likelihood ratio tests confirm that multilevel logistic regression models had better fit than single-level logistic regression models with p-value <0.001. This is not reported in the paper because we feel the scientific reasoning supersedes statistical testing of comparison between models.***

In response, we have improved the rationale for multilevel modelling in the fourth paragraph in Introduction section:

> Notably, the majority of multilevel studies on tobacco use to date investigate associations between specific area-level exposures and tobacco use (the specific contextual effect). Such models are used simply as an extension of single-level regression models enabling them to handle group-level variables as exposures and covariates. Variation in tobacco use across contexts (general contextual effects) can also be examined using multilevel models. Yet, this aspect of multilevel analysis has been underutilized in research to date. [24 25] Using this approach, we can describe the extent of geographic inequalities in tobacco use drawing attention to underlying contextual drivers unaddressed through individually directed interventions. [26-29] This is important information. Tobacco control interventions targeting specific area-level exposures will only be effective if areas share significant inter-individual variation in tobacco use. [24 25]

Comment 4: Not using the policy and economic variables puts a serious limitation on the study as the tobacco use in any area will be greatly affected by the level of tobacco control policy implementation and the economic capacity of population in that area. Even if in the author's estimation this may result in 'atomistic fallacy', it would be good to include it for the readers to get a better understanding of the actual scenario as the information is based on the same data source.

Response: We have a limited range of available policy and economic variables. The variables referred to by R2 are not asked of non-tobacco users, and as such, their inclusion would substantially reduce the analytical sample size. More generally, we do not want to weaken the design of the study by attributing individual-level perceptions of policy availability, implementation and expenditures on tobacco use to geographic/administrative small areas (atomistic fallacy).

Comment 5: While the main purpose of paper is to determine the spatial variation in tobacco use at ward and village level, it is not clear why states have not been included in the models which is likely to effect the analyses significantly.

Response: Thank you for this suggestion. We have now included states within the multilevel multivariable logistic regression models as a covariate. We found that including states substantially explained area level variation in any tobacco use as well as in different types of tobacco use. We have modified the methods, results and discussion section to reflect this.

As an aside, one assumption of multilevel modelling is that both individuals and clusters (in this case ward/village) must be randomly selected from their source populations. Given that there is no randomization at the state level, states were not included as random intercepts within the main analyses. This is debated (http://www.bristol.ac.uk/cmm/learning/videos/random-intercepts.html.), therefore, we have performed a sensitivity analyses using states as random intercepts fitting three-level hierarchical model: individuals nested within citywards/villages nested within states. We found high correlation in overall tobacco use and different types of tobacco use within states as well as within same citywards/villages in same states. We have added this table as a supplementary appendix.

Comment 6: Page 2 line 18 and pg5 line 58: Total sample size of GATS-2 was 74,037 individuals from 30 states and two union territories and not 31 states and union territories.

Response: Thank you. Corrected.

Comment 7: (Results) It is not clear how the results have been controlled for the impact of state, district and local level tobacco control policy and implementation of COTPA.
Response: We have now included states as a covariate within the model (see response to Comment 5). Given that small areas cannot be identified within the released GATS-2 dataset, we cannot include variables for state-, district- and local-level tobacco control policy and implementation of COTPA. We now include this as a research implication:

> Given the findings from our study, future GATS surveys should consider the opportunities to comprehensively study both individual- and area-level determinants of tobacco use within India and in other LMICs. First, it would be helpful if wards and villages were identifiable in future versions of GATS so that researchers and policymakers can link in area-level covariates (social, policy, economic and physical environment) to examine their effects on tobacco use. Second, it would be useful if the administrative levels at which tobacco related policies are implemented were recorded, allowing examining of variation in tobacco use across multiple levels of geographical hierarchy. This would further help policymakers compare clusters from an intervention perspective. Finally, identification of city wards and villages would also allow linking data to relevant area-level social, demographic, economic and policy variables increasing the ability to simultaneously examine area- and individual-level determinants of tobacco use.

Comment 8: It would be more helpful if the geographical variations are compared and presented for GATS-1 vs GATS-2. In addition if this is further looked from an NTCP v. non-NTCP districts/ward/village that will also help in arriving at more specific policy recommendations for strengthening NTCP at local level.
Response: This is related to the previous point. As districts/ward/villages are non-identifiable within the GATS dataset in either survey this means that we cannot match and compare areas between GATS-1 and 2.

Comment 9: It is suggested to provide full model details with all independent variables with corresponding OR, P-value and 95%CI for the benefit of readers.
Response: We have now provided this information as a supplementary appendix.

Comment 10: Table 1 (descriptive statistics) does not include urban-rural but mentioned in remaining table footnotes. Please clarify whether subsequent multilevel models are adjusted for urban-rural? As mentioned previously, it is much better to provide the full details of all tables rather than just writing in the footnote for better clarity, in case of space limitation it can be included as online supplement.
Response: Thank you for spotting this. Area of residence was included across all multilevel models in Model C. We have corrected this in footnotes and included the estimate within Table 1.

Comment 11: Pg 10 line 6: Not able to follow the mentioned urban rural variance in the table-2
Response: We've now clarified the description of this in the text. The text reads:

> We did this in three stages. First, we fitted a single-level logistic regression model with tobacco use as the outcome and included individual-level covariates (age, sex, education, household wealth and occupation) (Model A). The ability of this model to classify tobacco use was quantified using the Area Under Curve (AUC). Next, we fitted a multilevel logistic regression model (Model B) for tobacco use that included the same individual-level covariates. In addition to quantifying the change in the AUC from Model A, MORs and ICCs were estimated from Model B to examine the general contextual effect of areas. Finally, we added area of residence and states in Model C as area-level covariates to examine any changes in AUC, MOR and ICCs.

We did not see any meaningful changes in the estimates of AUC, MOR and ICCs upon just inclusion of area of residence as a covariate in Tables 2 and 3. However, updated model C reflects the role of state in observed variations in tobacco use.

Comment 12: Pg 10 line 10: what was the median odds ratio for rural areas?

Response: We would like to clarify that the median odds ratios were not stratified by area of residence. Therefore, it is not possible to calculate median odds ratio for rural areas.

Comment 13 (Discussion)**:** It would be helpful for the readers if the discussion adds on what are the possible reasons for higher clustering of tobacco use at area level? And also the possible reasons for higher clustering of dual use and SLT use?
Response: We have added the following text to discussion section:

> Our findings indicate much higher clustering of tobacco use at the area level than has previously been reported, suggesting that local area contexts and contextual determinants are highly relevant in India. Such variations, we speculate in the absence of data and available literature,[10-22] may be due to differences in the availability and implementation of tobacco control policies, social environment (deprivation, area-level mean income, area-level income inequality, social capital) and shared cultural and social norms regarding tobacco use among people within an area.

Comment 14: Whether data from GATS-1 and GATS-2 are unable to explain the factors responsible for the high variations of tobacco use at individual level or area level or both? How this limitation can be overcome? Please give some examples of area-level determinants that is being suggested to be included for future GATS surveys.
Response: We have added the following text in the discussion:

> First, it would be helpful if wards and villages were identifiable in future versions of GATS so that researchers and policymakers can link in area-level covariates (social, policy, economic and physical environment) to examine their effects on tobacco use. Second, it would be useful if the administrative levels at which tobacco related policies are implemented were recorded, allowing examining of variation in tobacco use across multiple levels of geographical hierarchy. This would further help policymakers compare clusters from an intervention perspective. Finally, identification of city wards and villages would also allow linking data to relevant area-level social, demographic, economic and policy variables increasing the ability to simultaneously examine area- and individual-level determinants of tobacco use.

Comment 15: Policy implications of the paper can be further discussed keeping in mind the structure of
governance at the national/state/district/local level and compliance with NTCP, COTPA and other tobacco control policies at these different levels.
Response: We have now further stressed on the need to consider local areas as important level of social organization to implement tobacco control interventions. Additionally, we have now emphasized the need to identify local areas to allow investigating social, policy, economic and physical environment relevant to tobacco use. Finally, identifying local areas will also help in prioritising areas with high levels of tobacco use.

Comment 16: "Comparison of GATS-2 and GATS-1 has highlighted changes in prevalence of tobacco use due to differential implementation of these measures." (Reference?)
Response: Thank you for pointing this out. We have added the following two references to support our claim:

1. ***Ahluwalia IB, Arrazola RA, Zhao L, et al. Tobacco Use and Tobacco-Related Behaviors - 11 Countries, 2008-2017. MMWR Morbidity and mortality weekly report 2019;68(41):928-933.***

2. ***Nazar GP, Chang KC, Srivastava S, et al. Impact of India's National Tobacco Control Programme on bidi and cigarette consumption: a difference-in-differences analysis. Tobacco control 2020;29(1):103-110.***

Comment 17: "States are also allowed to develop context specific information, education and communication resources to match the local needs." (Refrence?)

Response: We have added the following two references:

1. ***Nazar GP, Chang KC, Srivastava S, et al. Impact of India's National Tobacco Control Programme on bidi and cigarette consumption: a difference-in-differences analysis. Tobacco control 2020;29(1):103-110.***
2. ***NHM. National Tobacco Control Programme (NTCP): National Health Mission, Ministry of Health and Family Welfare, Government of India; 2019 [Available from: https://nhm.gov.in/index1.php?lang=1&level=2&sublinkid=1052&lid=607 accessed 6/3/2020 2020.***

Comment 18: It will be helpful for the readers to have some idea of the current design and reach of the National Tobacco Control Programme at district, block and village level?
Response: Thank you for this suggestion. While we recognize that India's National Tobacco Control Programme (NTCP) is the key national tobacco control program, it is out of scope for this paper to comprehensively canvass the design of the program. We have added the following text to incorporate your suggestion:

> NTCP is rolled out in 612 districts across 36 states/union territories in India and has a three-tier structure: National-, State- and District Tobacco Control Cell. District Tobacco Control Cells are established to train key stakeholders; information, education and communication activities; school programmes; monitor tobacco control laws; strengthen cessation facilities and co-ordinate tobacco control activities with Panchayati Raj (traditional local self-governance).[42] High local-area variations in tobacco use reported in our study imply extending this structure more locally to city-wards and villages to maximise public health benefits.

Comment 19: Table 1 on page 18 has two categories of dual use which is confusing and non-confirming to the definition of dual use in the paper
Response: We have now merged the two groups.

Reviewer # 3

Comment 1: The authors present an interesting analysis of the impact of geographic location on the prevalence of tobacco products in India. I think that the manuscript adds to the literature and provides important information. Yet, I feel that the manuscript would benefit from revision in several areas.
Response: Many thanks for this and comment and for your other suggestions about how to improve the manuscript.

Comment 2: As currently written and analyzed, it is hard to understand how the underlying prevalence data is reflected in the statistical models. For this reason, I think the authors should provide the prevalence of each tobacco product in each geographic location included in the analysis for readers. This could be done via a table and a map. For example, how does the prevalence of each type of tobacco & dual use vary across the included geographic areas?
Response: Thank you for pointing this out. The clusters (villages and city wards) in GATS are non-identifiable. They are only represented with anonymous cluster IDs. As such, we cannot take up this excellent suggestion of mapping them or examining areas in detail where there is high prevalence. But we have added plots for the prevalence and 95% confidence intervals for each cluster of any tobacco use, cigarette smoking, bidi smoking, smokeless tobacco use and dual use for all clusters in the supplementary appendix. We have done this as a caterpillar plot, ordered from lowest to highest prevalence. We think this shows the information you are interested in.

Comment 3: Then, I think it would be useful for the authors to further justify and explain the advantages of their approach versus simply examining the prevalence of each tobacco product in each geographic location using simpler methods. And further, how is the information gained from these analysis help

make public health decisions. For example, one approach might be to target areas with high prevalence—or alternatively, areas with a large # of users. How would the information gained by the analyses described herein contribute to such decisions? Or would it?
Response: We have now described in paragraph 4 of the background how the information generated from investigating extent of geographic variation in tobacco use and its determinants will contribute to public health decision:

> Notably, the majority of multilevel studies on tobacco use to date investigate associations between specific area-level exposures and tobacco use (the specific contextual effect). Such models are used simply as an extension of single-level regression models enabling them to handle group-level variables as exposures and covariates. Variation in tobacco use across contexts (general contextual effects) can also be examined using multilevel models. Yet, this aspect of multilevel analysis has been underutilized in research to date. [24 25] Using this approach, it is possible to describe the extent of geographic inequalities in tobacco use drawing attention to underlying contextual drivers unaddressed through individually directed interventions. [26-29] This is important information. Tobacco control interventions targeting specific area-level exposures will only be effective if areas share significant inter-individual variation in tobacco use. [24 25]

Comment 4: The abstract should spell out all abbreviations, such as MOR, and further should detail the methods used.
Response: Done.

Comment 5: Results, page 10, top paragraph: This conclusion seems overstated. As elsewhere, tobacco use typically begins in India in youth—thus it seems very unlikely to me that a middle-aged individual who doesn't use tobacco would begin because they moved to a new area. Perhaps geographic differences are most correlated in rates of initiation and cessation?
Response: We have corrected this. The text now reads:

> <u>These results suggest that the median odds of tobacco use are more than double for two individuals with same covariates when comparing the one from city-ward or village with high tobacco use to the other from a city-ward or village with low tobacco use.</u>

Comment 6: The authors suggest that future GATS surveys should add questions in order to assess area-level determinants (page 12). I think this statement would benefit from clearly describing what sorts of questions should be added.
Response: We have added text to describe the necessary information for assessment of area-level determinants:

> <u>First, it would be helpful if wards and villages were identifiable in future versions of GATS so that researchers and policymakers can link in area-level covariates (social, policy, economic and physical environment) to examine their effects on tobacco use. Second, it would be useful if the administrative levels at which tobacco related policies are implemented were recorded, allowing examining of variation in tobacco use across multiple levels of geographical hierarchy. This would further help policymakers compare clusters from an intervention perspective. Finally, identification of city wards and villages would also allow linking data to relevant area-level social, demographic, economic and policy variables increasing the ability to simultaneously examine area- and individual-level determinants of tobacco use.</u>

VERSION 2 – REVIEW

REVIEWER	Amit Yadav University of California San Francisco, San Francisco, USA
REVIEW RETURNED	01-Apr-2020

GENERAL COMMENTS	Thank you for taking the inputs and revising the manuscript and making it more easy and accessible. I have the following minor observations on the revised version. I look forward to seeing this important work in print very soon! 1. Numbers from GATS in abstract and methods sections are different 2. In table 1-adding bidi+cigarette smokers to dual-use are not commensurate with the dual-use definition in the paper. instead of the cig only, bidi only and bidi+cig they all can be clubbed as smoking population. in case that is problem in analysis bidi+cig can be kept separate category but not calling them dual-use. 3. Figure-1 is missing from the revised draft

REVIEWER	Neal Freedman National Cancer Institute, Division of Cancer Epidemiology and Genetics
REVIEW RETURNED	13-Apr-2020

GENERAL COMMENTS	I think that the authors have done an excellent job with their revisions and I have no further comments.

VERSION 2 – AUTHOR RESPONSE

Reviewer# 2

Comment 1: Thank you for taking the inputs and revising the manuscript and making it more easy and accessible.
Response: Thank you.

Comment 2: Numbers from GATS in abstract and methods sections are different.
Response: Thank you. We have corrected the numbers in methods section.

Comment 3: In table 1-adding bidi+cigarette smokers to dual-use are not commensurate with the dual-use definition in the paper. instead of the cig only, bidi only and bidi+cig they all can be clubbed as smoking population. in case that is problem in analysis bidi+cig can be kept separate category but not calling them dual-use.
Response: We have revised Table 1. For descriptive purposes bidi+cigarettes are now kept separate.

Comment 4: Figure-1 is missing from the revised draft.
Response: Figure-1 has been attached separately as required by the journal.