BMJ Open

Geographic variation in tobacco use in India: A population based multi-level study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-033178
Article Type:	Original research
Date Submitted by the Author:	24-Jul-2019
Complete List of Authors:	Singh, Ankur; University of Melbourne, Centre for Health Equity, Melbourne School of Population & Global Health Arora, Monika; Public Health Foundation of India, Health Promotion and Tobacco Control Bentley, Rebecca; University of Melbourne, Centre for Health Equity, Melbourne School of Population & Global Health; University of Melbourne, Centre for Epidemiology and Biostatistics, Melbourne School of Population & Global Health Spittal, Matthew; The University of Melbourne, Do, Loc; The University of Adelaide, Australian Research Centre for Population Oral Health (ARCPOH), Adelaide Dental School Grills, Nathan; The University of Melbourne, Nossal Institute for Global Health English, Dallas; University of Melbourne, Melbourne School of Population and Global Health
Keywords:	Tobacco, India, EPIDEMIOLOGY, PUBLIC HEALTH, SOCIAL MEDICINE

Geographic variation in tobacco use in India: A population based multi-level study

List of authors: Ankur Singh (Ph.D.), Monika Arora (Ph.D.), Rebecca Bentley (Ph.D.), Matthew J Spittal (Ph.D.), Loc G Do (Ph.D.), Nathan Grills (Ph.D.), Dallas R English (Ph.D)

Corresponding author:

Dr. Ankur Singh

Centre for Health Equity, Melbourne School of Population and Global Health,

The University of Melbourne,

Level 3, 207 Bouverie Street,

Melbourne, Victoria 3010, Australia

Tel: +61 3 8344 9256 Email: ankur.singh@unimelb.edu.au

Journal: BMJ Open

Publication category: Original Research

Running head: Area-level variations in tobacco use in India

Word count (excluding abstract, references, tables, and figures): 2655 words

Number of references: 39

Conflict of interest: None Declared

Abstract

Objective: This study aims to quantify the extent to which people's use of tobacco products varies by local areas (city-ward/village) across India and the variation in this clustering by tobacco product.

Design: Cross-sectional study

Setting and participants: Data on a total of 74,037 adults across 2,547 city wards and villages was available for analysis from 31 states and union territories in India.

Primary and secondary outcome measures: We included as primary outcomes self-reported any tobacco use, current cigarette smoking, current bidi smoking, current smokeless tobacco use and a derived variable for dual use describing respondents who engaged in both smoking and smokeless tobacco use.

Results: The median risk of an individual using tobacco was 2.42 times greater if a person hypothetically moved from an area of low to high risk of tobacco use (95% CI 2.34 – 2.51). Area-level partitioning of variation differed by tobacco product used. MORs ranged from 3.14 for cigarette smoking to 4.39 for dual use.

Conclusion: Tobacco use is highly geographically clustered in India. To be effective in India policy interventions must, therefore, account for the influence of specific local contextual factors on adult tobacco use. Where people live in India influences their use of tobacco, and this association may be greater than has been observed in other settings. Tailoring tobacco control policies for local areas in India may, therefore, provide substantial public health benefits.

Strengths and limitations of this study:

- This is the first study from LMICs that has studied variation in tobacco use at local area level using a nationwide representative data.

- By using different measures (ICC, MOR and AUC) we not only inform the extent of variation but we comprehensively examine the degree of clustering, the heterogeneity in outcomes among areas as well as the ability of local areas to classify individuals according to tobacco use.
- We did not incorporate policy and economic variables related to tobacco use available in the GATS 2016-17 in our analysis. The variables used were non-ecologic and could potentially lead to the atomistic fallacy by falsely attributing individual-level measures to areas.

Introduction

Four fifths of the world's current smokers reside in low- and middle-income countries (LMICs), creating enormous societal and public health challenges. [1] The number of deaths from tobacco-related causes and loss of productivity is rapidly increasing in these, often resource poor, settings. [2 3]

The latest Indian Global Adult Tobacco Survey found that nearly 30% of all Indian adults use tobacco. [4] Additionally, the widespread use of smokeless tobacco presents a complex challenge for health systems and tobacco control because of its strong association with oral cancerous and pre-cancerous lesions. [5] Despite a nation-wide smokeless tobacco ban implemented in 2013-14, 20% of all tobacco users are smokeless tobacco users. [4] Added to this, the burden of tobacco use in India is disproportionally high among people who are socially disadvantaged. [6-8]

There is consistent evidence that local social and policy contexts shape patterns of tobacco use. [9] Multilevel studies (that simultaneously examine individual- and group-level determinants of health) from The Netherlands, Australia, South Africa, Mexico, Scotland, India, the USA and the UK suggest evidence of an association between area-level context (such as social disadvantage and local policy environments) and smoking.[10-22] For example, a study of Indian high school students from Mumbai reported the density of tobacco vendors around schools was associated with increased tobacco use by students. [23]

Notably,the majority of multilevel studies on tobacco use to date investigate associations between specific exposures and tobacco use (the specific contextual effect). The variation in tobacco use across different contexts (general contextual effect), including India, has not been quantified and described. [24 25] This is important for several reasons. First, describing the extent of geographic inequalities in tobacco use draws attention to underlying contextual drivers that cannot be addressed through individually directed interventions. [26-29] Second, tobacco control interventions targeting specific area-level exposures will only be effective if areas share significant inter-individual variation in tobacco use. [24 25] Finally, due to its impact on effective sample sizes, a small general contextual effect can lead to detection of small but statistically significant specific contextual effects. [25] Therefore, the observed association between specific area-level exposures and tobacco use (specific contextual effect) may be spurious and lead to targeting non-relevant determinants when general contextual effect are ignored.

To redress this important gap in evidence, this study aims to quantify the extent to which people's use of tobacco products varies by local areas (city-ward/village) across India and the variation in this clustering by tobacco product.

Methods

Study population

Data on tobacco use in India was obtained from the Global Adult Tobacco Survey (GATS 2) conducted in 2016 and 2017. GATS 2 is a multi-country household tobacco prevalence survey designed to support implementation of tobacco control within study countries.[4] Participants eligible for the survey were non-institutionalised individuals aged 15 years and older. The survey applied a multistage sampling procedure with different sampling hierarchies for urban and rural areas. For urban areas, city wards were the primary sampling unit from which census enumeration blocks, and then households, were selected. In rural areas, the primary sampling units were villages, from which households were selected. A total of 74,037 adults across 2,547 city wards and villages were available for analysis from 31 states and union territories in India. The response rate was 93%. [4]

Data collection

GATS-2 collected data using household and individual questionnaires developed in English and translated into 19 regional languages. The interviewer-administered questionnaires collected data on demographic characteristics, tobacco smoking, smokeless tobacco use, second hand smoke, socioeconomic position, media and knowledge, attitude and perceptions related to tobacco use. More details on sampling procedures and methods of data collection are published elsewhere.[4 30 31]

Outcomes

We included as primary outcomes self-reported any tobacco use, current cigarette smoking, current bidi smoking, current smokeless tobacco use and a derived variable for dual use describing respondents who engaged in both smoking and smokeless tobacco use. Participants were asked 'On average, how many of the following products do you currently smoke each day?.[4 30 31] We categorized those who reported smoking one or more manufactured/rolled tobacco in paper/leaf as current cigarette smokers. Similarly, we identified those who reported smoking one or more bidi as current bidi smokers. Regarding smokeless tobacco use, participants were asked 'Do you currently use smokeless tobacco on a daily basis, less than daily, or not at all? '.[4 30 31] We recorded those answering 'daily' or 'less than daily' as yes for current smokeless tobacco use. Those identified to be both current smokers (cigarette or bidi) and current smokeless tobacco users were identified as dual users. Therefore, we created five binary variables including any tobacco use, current cigarette smokers, current bidi smokers, current smokeless tobacco users and dual users.

Geographic level of aggregation (local areas)

Individuals from urban areas were clustered within city-wards and those in rural areas were clustered within villages. In urban areas, city wards are the units for local government

operations in India, responsible for essential community services including healthcare, education, housing, transport and so on. [32] In rural areas, villages make up the boundary for local panchayat (traditional local self-governance). [32]

Covariates

To account for compositional differences in populations within area-level clusters, we included individual-level demographic characteristics: age (as a continuous variable), sex and socioeconomic position (education (no formal education/less than primary/primary/ secondary or more), occupation (unemployed/labourer/housewife, retired, student/ self-employed/ private/ government) and household-level wealth (quintiles, 1 = lowest, 5 = highest)) as covariates in the multilevel regression models. These variables were selected based on a previous study. [7]

Statistical Analysis

We performed the statistical analyses using Stata 15.0 (Statacorp, College Station, TX, USA). We used survey commands to account for the complex survey design and to perform the weighted descriptive analysis. We fitted multilevel logistic regression models with random intercepts for local areas and fixed slopes with individuals nested in city wards or villages respectively. Multilevel models operationalise studying population-level variations in health outcomes by examining the extent of clustering in health outcomes that exists at the group or contextual level. [24 33-37] Using intra-class correlation coefficients (ICC) and median odds ratios (MOR), we decomposed the variance in health outcomes at different levels of social organisation. The ICC is expressed as a percentage and is interpreted in these analyses as the share of inter-individual variation in health outcome that exists at the group level. For example, an ICC of 8% at the village level means that of all the individual-level variation in tobacco use among rural areas, 8% is attributed to the village level. The higher the individual correlation in health outcomes within a context, the more relevant is the context for understanding individual differences in the health outcome. [25] We estimated the MOR, which describes the area-level variance as an odds ratio, as the median value of the distribution of odds ratios obtained when two individuals with the same covariate values are picked from two different areas, comparing the one from the higher prevalence area to the one from the area with lower prevalence. [24 36 38] In the absence of any area-level variation, the MOR is equal to one. We estimated both MORs and ICCs for binary outcomes as the partition of variance between different levels does not have the intuitive interpretation of the linear model. [24 38] We estimated ICCs and MORs from intercept only models to examine presence of clustering, and heterogeneity between areas, in the outcomes of tobacco use.

We also applied an alternate method to examine the relevance of area-level contexts for tobacco use by comparing discriminatory accuracies obtained from fitted single-level and multi-level logistic regression models. The area under the receiver operating characteristic curve (AU-ROC) was constructed by plotting the true positive fraction (TPF, sensitivity) against the false positive fraction (FPF, 1 – specificity). It measures the ability of the model to classify individuals with and without the outcome and takes a value between 0.5 and 1.0 where 1.0 is perfect discrimination and 0.5 where covariates have no predictive power. [24]

First, we fitted a single-level logistic regression model with tobacco use as 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 in Model C as an area-level covariate to examine any changes in AUC, MOR and ICCs.

We assessed goodness of fit by estimating the changes in the Deviance Information Criterion (DIC). All models were fitted separately for each type of tobacco use (cigarette smoking, bidi smoking, smokeless tobacco use and dual use) to determine any differences in variations in tobacco use according to different types of tobacco use.

Patient and Public Involvement

No patients or public were involved in this study.

Results

We analysed data for 73,954 individuals (99.9%) of the 74,037 survey participants. We did not analyse data on 83 participants due to missing data on covariates. Table 1 shows descriptive characteristics of the sample according to residence status. 28% of adults used tobacco products. The prevalence of smokeless tobacco use was 18.6% (Table 1).

Intercept only models (null models with no covariate adjustment) estimated 22% (95% CI: 20, 24) of any tobacco use was clustered at the city-ward/village level. Cigarette smoking was clustered 31%, bidi smoking at 28%, dual use at 40% and smokeless tobacco at 36% respectively (estimates not reported in the tables). For each outcome, the AUC increased when multilevel logistic regression models were fitted (Model B) as compared to single-level logistic regression models (Model A) implying the presence of a general contextual effect (Figure 1). Changes in AUC were highest for smokeless tobacco use 11%, compared to 2% for cigarette smoking (Table 2 and 3).

After including all individual-level covariates, the proportion of variance attributable to the areas remained at 21% (95% CI: 20, 22) for urban areas and increased to 20% (95% CI: 19, 22) for rural areas (Table 2). Correspondingly, the median odds ratio for urban areas was 2.42 (95% CI: 2.35, 2.52). These results suggest that if an individual moved to a city-ward or village with high tobacco use from a city-ward or village with low tobacco use, their median odds of tobacco use would more than double. No substantial variation in estimates were observed when area of residence was included in Model C compared to only individual-level covariates in Model B (Table 2).

The decrease in DIC values between the single-level models and multilevel models including covariates suggested better model fit (Table 2).

Among the different types of tobacco use, the highest ICC (42%; 95% CI: 39, 46) and MOR (4.39; 95% CI: 3.97, 4.82) were for dual use and the lowest for cigarette smoking ((ICC: 30%; 95% CI: 27, 34), (MOR: 3.14; 95% CI: 2.82, 3.46)) (Table 3).

Discussion

We found substantial variation in tobacco use across local areas in India. Individual-level social and demographic characteristics were not able to explain the high area-level variations in tobacco. The degree of area-level variation in tobacco use differed according to the types of tobacco product. Dual use (smoking and smokeless) had the highest geographic clustering.

Strengths and Limitations

This study had several strengths and some limitations. To the best of our knowledge this is the first study from LMICs that has studied variation in tobacco use at local area level using a nationwide representative data. [4] By using different measures (ICC, MOR and AUC) we not only inform the extent of variation but we comprehensively examine the degree of clustering, the heterogeneity in outcomes among areas as well as the ability of local areas to classify individuals according to tobacco use. [24 33-36] This study also has limitations. We did not incorporate policy and economic variables related to tobacco use available in the GATS 2016-17 in our analysis. This decision was underpinned by the reasoning that the policy and economic variables were respondents perceptions of policy and individual expenditures on tobacco use rather than availability and implementation of policies in local areas. Therefore, the variables used were non-ecologic and could potentially lead to the atomistic fallacy by falsely attributing individual-level measures to areas. [37]

Discussion in context of current evidence

Our findings of high variations in tobacco use among local areas is new. A multilevel study on societal determinants of tobacco use from Scotland found no evidence of clustering in tobacco use at the area level. [18] Other multilevel studies have not presented measures of variance, which limits comparisons.[10-22] 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 may be due to differences in the availability and implementation of tobacco control policies.

Tobacco specific variations in the values of ICC and MOR highlight potential differences in the relevance of contexts by type of tobacco product used. Evidence from other studies suggests that while wealthier and more educated individuals have higher odds of cigarette smoking than their disadvantaged counterparts, disadvantaged individuals have higher odds of bidi smoking and smokeless tobacco use. [6 7] Our study highlights the presence of both individual and geographic socioeconomic inequalities in tobacco use by product. For example, we observed a higher effect of individual social and demographic characteristics in smokeless tobacco use when compared to cigarette smoking and bidi smoking for contextual effects (change in ICC from 36% in null model to 30% in adjusted model).

Research and policy implications

Given the role of contexts in shaping individual health behaviours, this study builds a framework for operationalizing a contextual thinking in tobacco control activities, particularly in LMICs where social norms and cultural aspects may differ from high-income countries. High general contextual effects of local areas for tobacco use necessitates a thorough examination of factors at the area-level that may be causally associated with individual tobacco use as well as those which can explain the high variations in tobacco use among local areas. This may only be possible if either data on individual-level tobacco use is linked with small area characteristics, or if future population-based surveys collect both area- and individual-level data relevant to tobacco use. 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. Furthermore, current findings build the platform for more robust population-based studies that collectively examine area- and individual-level determinants of tobacco use in India and other LMICs.

This study has several policy implications. Our findings confirm that context plays an important role in determining use of tobacco. India's Cigarettes and other Tobacco Products Act (COTPA) is a national law, which is in line with World Health Organization's Framework Convention on Tobacco Control. States at sub-national level are responsible for implementing various tobacco control policy measures under COTPA. Comparison of GATS-2 and GATS-1 has highlighted changes in prevalence of tobacco use due to differential implementation of these measures. States are also allowed to develop context specific information, education and communication resources to match the local needs.

Therefore, health promotion and tobacco control interventions must be designed for contexts and applied contextually rather than being individually oriented.[9 39] There is need to enhance National Tobacco Control Program's implementation at district, village and block level as well. Finally, our use of the multilevel approach in this study advances a 'proportionate universalism' approach. Tobacco control intereventions applied nationally should be scaled according to local area characteristics to reduce geographic inequalities.

Conclusion

Where people live in India influences their use of tobacco, and this association may be greater than has been observed in other settings. Tailoring tobacco control policies for local areas in India may, therefore, provide substantial public health benefits.

References

1. WHO. Tobacco: Fact sheet Geneva: World Health Organization; 2018 [Available from: https://www.who.int/en/news-room/fact-sheets/detail/tobacco accessed 10 December 2018.
2. Hipple Walters B, Petrea I, Lando H. Tobacco Control in Low- and Middle-Income Countries: Changing the Present to Help the Future. *Journal of Smoking Cessation* 2018:1-2. doi: 10.1017/jsc.2018.4 [published Online First: 03/14]
3. Gilmore AB, Fooks G, Drope J, et al. Exposing and addressing tobacco industry conduct in low-income and middle-income countries. *The Lancet* 2015;385(9972):1029-43. doi: 10.1016/S0140-6736(15)60312-9
4. Tata Institute of Social Sciences (TISS) MaMoHaFW, Government of India. Global Adult Tobacco Survey GATS 2 India 2016-17. New Delhi: Tata Institute of Social Sciences (TISS), Mumbai and Ministry of Health and Family Welfare, Government of India, 2018.
5. Gupta PC, Arora M, Sinha D, et al. Smokeless tobacco and public health in India. *Ministry of Health & Family Welfare, Government of India; New Delhi* 2016
6. Corsi DJ, Subramanian SV. Divergent socio-economic gradients in smoking by type of tobacco use in India. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease* 2014;18(1):122-4. doi: 10.5588/ijtld.13.0246 [published Online First: 2013/12/25]
7. Singh A, Arora M, English DR, et al. Socioeconomic Gradients in Different Types of Tobacco Use in India: Evidence from Global Adult Tobacco Survey 2009-10. *Biomed Res Int* 2015;2015:837804. doi: 10.1155/2015/837804 [published Online First: 2015/08/15]
8. Bhan N, Srivastava S, Agrawal S, et al. Are socioeconomic disparities in tobacco consumption increasing in India? A repeated cross-sectional multilevel analysis. *BMJ open* 2012;2(5) doi: 10.1136/bmjopen-2012-001348 [published Online First: 2012/10/02]
9. Link BG, Phelan J. Social Conditions as Fundamental Causes of Disease. *Journal of Health and Social Behavior* 1995;35:80-94. doi: Doi 10.2307/2626958
10. Rachele JN, Wood L, Nathan A, et al. Neighbourhood disadvantage and smoking: Examining the role of neighbourhood-level psychosocial characteristics. *Health Place* 2016;40:98-105. doi: 10.1016/j.healthplace.2016.04.012 [published Online First: 2016/05/27]
11. Ellaway A, Macintyre S. Are perceived neighbourhood problems associated with the likelihood of smoking? *Journal of epidemiology and community health* 2009;63(1):78-80. doi: 10.1136/jech.2007.068767 [published Online First: 2008/12/18]
12. Morris T, Manley D, Van Ham M. Context or composition: How does neighbourhood deprivation impact upon adolescent smoking behaviour? *PloS one* 2018;13(2):e0192566. doi: 10.1371/journal.pone.0192566 [published Online First: 2018/02/09]
13. Mayne SL, Auchincloss AH, Moore KA, et al. Cross-sectional and longitudinal associations of neighbourhood social environment and smoking behaviour: the multiethnic study of atherosclerosis. *Journal of epidemiology and community health* 2017;71(4):396-403. doi: 10.1136/jech-2016-207990 [published Online First: 2016/11/26]
14. Chuang YC, Cubbin C, Ahn D, et al. Effects of neighbourhood socioeconomic status and convenience store concentration on individual level smoking. *Journal of*

epidemiology and community health 2005;59(7):568-73. doi: 10.1136/jech.2004.029041 [published Online First: 2005/06/21]
15. Fleischer NL, Lozano P, Arillo Santillan E, et al. The impact of neighbourhood violence and social cohesion on smoking behaviours among a cohort of smokers in Mexico. *Journal of epidemiology and community health* 2015;69(11):1083-90. doi: 10.1136/jech-2014-205115 [published Online First: 2015/06/06]
16. Turrell G, Hewitt BA, Miller SA. The influence of neighbourhood disadvantage on smoking cessation and its contribution to inequalities in smoking status. *Drug Alcohol Rev* 2012;31(5):645-52. doi: 10.1111/j.1465-3362.2012.00452.x [published Online First: 2012/04/18]
17. van Lenthe FJ, Mackenbach JP. Neighbourhood and individual socioeconomic inequalities in smoking: the role of physical neighbourhood stressors. *Journal of epidemiology and community health* 2006;60(8):699-705. doi: 10.1136/jech.2005.043851 [published Online First: 2006/07/15]
18. Shareck M, Ellaway A. Neighbourhood crime and smoking: the role of objective and perceived crime measures. *BMC public health* 2011;11:930. doi: 10.1186/1471-2458-11-930 [published Online First: 2011/12/16]
19. Fleischer NL, Thrasher JF, Saenz de Miera Juarez B, et al. Neighbourhood deprivation and smoking and quit behaviour among smokers in Mexico: findings from the ITC Mexico Survey. *Tobacco control* 2015;24 Suppl 3:iii56-iii63. doi: 10.1136/tobaccocontrol-2013-051495 [published Online First: 2014/08/30]
20. Lau YK, Tam J, Fleischer NL, et al. Neighbourhood deprivation, smoking, and race in South Africa: A cross-sectional analysis. *Prev Med Rep* 2018;11:202-08. doi: 10.1016/j.pmedr.2018.07.001 [published Online First: 2018/07/13]
21. Timmermans EJ, Veldhuizen EM, Snijder MB, et al. Neighbourhood safety and smoking in population subgroups: The HELIUS study. *Preventive medicine* 2018;112:111-18. doi: 10.1016/j.ypmed.2018.04.012 [published Online First: 2018/04/15]
22. Reijneveld SA. Neighbourhood socioeconomic context and self reported health and smoking: a secondary analysis of data on seven cities. *Journal of epidemiology and community health* 2002;56(12):935-42. [published Online First: 2002/12/04]
23. Mistry R, Pednekar M, Pimple S, et al. Banning tobacco sales and advertisements near educational institutions may reduce students' tobacco use risk: evidence from Mumbai, India. *Tobacco control* 2015;24(e1):e100-7. doi: 10.1136/tobaccocontrol-2012-050819 [published Online First: 2013/08/21]
24. Merlo J, Wagner P, Ghith N, et al. An Original Stepwise Multilevel Logistic Regression Analysis of Discriminatory Accuracy: The Case of Neighbourhoods and Health. *PloS one* 2016;11(4):e0153778. doi: 10.1371/journal.pone.0153778 [published Online First: 2016/04/28]
25. Merlo J, Wagner P, Austin PC, et al. General and specific contextual effects in multilevel regression analyses and their paradoxical relationship: A conceptual tutorial. *SSM - Population Health* 2018;5:33-37. doi: 10.1016/j.ssmph.2018.05.006
26. Diez Roux AV. The study of group-level factors in epidemiology: rethinking variables, study designs, and analytical approaches. *Epidemiologic reviews* 2004;26(1):104-11. doi: 10.1093/epirev/mxh006 [published Online First: 2004/07/06]
27. Merlo J. Multilevel analytical approaches in social epidemiology: measures of health variation compared with traditional measures of association. *Journal of epidemiology and community health* 2003;57(8):550-2. [published Online First: 2003/07/29]

28. Merlo J, Chaix B, Yang M, et al. A brief conceptual tutorial on multilevel analysis in social epidemiology: interpreting neighbourhood differences and the effect of neighbourhood characteristics on individual health. *Journal of epidemiology and community health* 2005;59(12):1022-8. doi: 10.1136/jech.2004.028035 [published Online First: 2005/11/16]
29. Merlo J, Viciana-Fernandez FJ, Ramiro-Farinas D, et al. Bringing the individual back to small-area variation studies: a multilevel analysis of all-cause mortality in Andalusia, Spain. *Soc Sci Med* 2012;75(8):1477-87. doi: 10.1016/j.socscimed.2012.06.004 [published Online First: 2012/07/17]
30. Palipudi KM, Morton J, Hsia J, et al. Methodology of the Global Adult Tobacco Survey - 2008-2010. *Global health promotion* 2016;23(2 Suppl):3-23. doi: 10.1177/1757975913499800 [published Online First: 2013/09/18]
31. GATS. Global Adult Tobacco Survey (GATS) India Report 2009-2010. : Ministry of Health and Family Welfare, India, 2010.
32. Office of the Registrar General & Census Commissioner I. Administrative Divisions New Delhi, India: Ministry of Home Affairs, Government of India; 2018 [Available from: http://censusindia.gov.in/Census_And_You/Administrative_division.aspx accessed 30/11/18 2018.
33. Subramanian SV. The relevance of multilevel statistical methods for identifying causal neighborhood effects. *Soc Sci Med* 2004;58(10):1961-7. doi: 10.1016/S0277-9536(03)00415-5 [published Online First: 2004/03/17]
34. Subramanian SV, Jones K, Kaddour A, et al. Revisiting Robinson: the perils of individualistic and ecologic fallacy. *International journal of epidemiology* 2009;38(2):342-60; author reply 70-3. doi: 10.1093/ije/dyn359 [published Online First: 2009/01/31]
35. Merlo J, Chaix B, Yang M, et al. A brief conceptual tutorial of multilevel analysis in social epidemiology: linking the statistical concept of clustering to the idea of contextual phenomenon. *Journal of epidemiology and community health* 2005;59(6):443-9. doi: 10.1136/jech.2004.023473 [published Online First: 2005/05/25]
36. Merlo J, Ohlsson H, Lynch KF, et al. Individual and collective bodies: using measures of variance and association in contextual epidemiology. *Journal of epidemiology and community health* 2009;63(12):1043-8. doi: 10.1136/jech.2009.088310 [published Online First: 2009/08/12]
37. Singh A, Harford J, Peres MA. Investigating societal determinants of oral health-Opportunities and challenges in multilevel studies. *Community dentistry and oral epidemiology* 2018;46(4):317-27. doi: 10.1111/cdoe.12369 [published Online First: 2018/02/21]
38. Merlo J, Chaix B, Ohlsson H, et al. A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *Journal of epidemiology and community health* 2006;60(4):290-7. doi: 10.1136/jech.2004.029454 [published Online First: 2006/03/16]
39. Phelan JC, Link BG, Tehranifar P. Social conditions as fundamental causes of health inequalities: theory, evidence, and policy implications. *J Health Soc Behav* 2010;51 Suppl(1_suppl):S28-40. doi: 10.1177/0022146510383498 [published Online First: 2010/12/22]

Footnotes

Author's contribution:

AS conceptualised the study, acquisition of data, analysed the data, interpretation of results, led the manuscript preparation and the submission process

MA contributed to interpretation of local policy implications of the results and drafts of the manuscript

RB and NG contributed by critical inputs on multiple draft of the manuscript and interpretation of results

MS and LD contributed by development of analytical framework, interpretation of results and revision of manuscript

DE contributed by reviewing early and advanced drafts of the manuscript, development of analysis plan and interpretation of results.

Funding statement: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests: None declared.

Declaration of Interest: Authors have no conflict of interest to declare.

Patient consent for publication: Not required.

Data sharing statement: Data are available upon reasonable request to the corresponding author. Email: ankur.singh@unimelb.edu.au

Ethics approval: No formal ethical approval was required.

Provenance and peer review: Not commissioned; externally peer reviewed.

Tables and figures

Table 1. Descriptive characteristics of the sample (n=73,954)

Variable	Categories	Percentage
Age (years)	15 to 30	41.7
	31 to 45	29.7
	46 to 60	17.7
	61 to 75	8.9
	76 and above	1.8
Sex	Male	51.1
	Female	48.9
Wealth	Poorer	23.4
	Poor	36.5
	Middle	15.0
	Rich	12.2
	Richer	12.9
Education	No formal education	26.4
	Less than primary	9.2
	Primary	28.2
	Secondary or more	36.2
Occupation	Unemployed	4.3
	Labour	21.2
	Housewife/ Retired/ Student	44.1
	Self	19.4
	Private	8.3
	Government job	2.7
Tobacco use	Non- user	72.2
	Cigarette smoking	1.3
	Bidi smoking	4.6
	Smokeless tobacco use	18.6
	Dual Use (Smokeless tobacco use + Smoking)	2.8
	Dual Use (Bidi + Cigarette)	0.5

Weighted percentages (Using Survey Weights)

Table 2. Multilevel logistic regression models for any tobacco use among Indian adults (n=73,954 individuals nested in 2547 city wards and villages).

	Model A Estimate 95% CI	Model B Estimate 95% CI	Model C Estimate 95% CI
AUC	0.79	0.86	0.86
AUC change		0.07	0
Variance		0.87	0.86
		(0.80, 0.94)	(0.79, 0.93)
ICC		21%	21%
		(20, 22)	(19, 22)
MOR		2.43	2.42
		(2.35, 2.52)	(2.34, 2.51)
DIC	71171.7	66619.6	66559.3
DIC change		-4552.1	-60.3

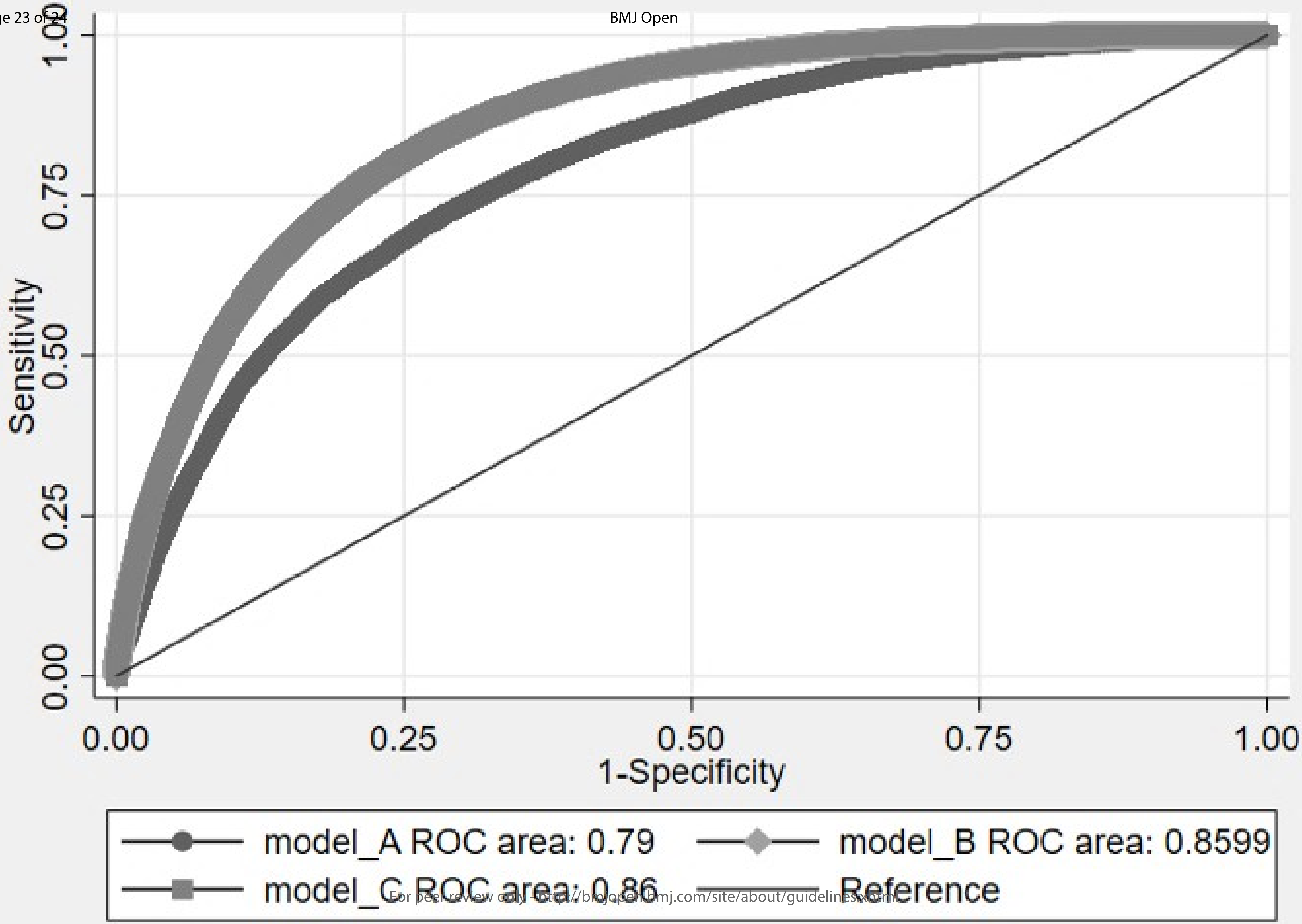
Model A: Single-level logistic regression model (Covariates included: age, sex, area of residence, education, occupation, wealth); Model B: Multi-level logistic regression model (Covariates included: age, sex, education, occupation, wealth); Model C: Multi-level logistic regression model (Covariates included: age, sex, education, occupation, wealth, area of residence). Reference group: No tobacco use.

Table 3. Multilevel logistic regression models for different types of tobacco use among Indian adults (n=73,954 individuals nested in 2547 city wards and villages).

	Model A Estimate 95% CI	Model B Estimate 95% CI	Model C Estimate 95% CI
Cigarettes (n=54,648)			
Variance		1.44 (1.19, 1.70)	1.44 (1.19, 1.70)
ICC (%)		30 (27, 34)	30 (27, 34)
MOR		3.14 (2.83, 3.46)	3.14 (2.82, 3.46)
DIC	10630.4	10175.0	10156.3
DIC change		-455.4	-18.7
AUC	0.68	0.70	0.70
AUC change		0.02	0
Bidi (n=56,814)			
Variance		1.53 (1.33, 1.72)	1.49 (1.30, 1.68)
ICC (%)		32 (29, 35)	31 (28,34)
MOR		3.25 (3.01, 3.49)	3.20 (2.97, 3.44)
DIC	18822.5	17680.8	17630.3
DIC change		-1141.7	-50.5
AUC	0.89	0.95	0.95
AUC change		0.06	0
SLT (n=66,089)			
Variance		1.46 (1.34, 1.59)	1.45 (1.33, 1.58)
ICC (%)		31 (29, 33)	31 (29, 32)
MOR		3.17 (3.01, 3.32)	3.16 (3.00, 3.31)
DIC	56207.3	51179.1	51129.0
DIC change		-5028.1	-50.2
AUC	0.76	0.87	0.87
AUC change		0.11	0
Dual use (n=55,522)			
Variance		2.41 (2.09, 2.72)	2.41 (2.09, 2.73)
ICC (%)		42 (39, 45)	42 (39, 46)
MOR		4.39 (3.96, 4.82)	4.39 (3.97, 4.82)
DIC	14335.7	12989.8	12977.1
DIC change		-1345.9	-12.7
AUC	0.88	0.96	0.96
AUC change		0.08	0

Model A: Single-level logistic regression model (Covariates included: age, sex, area of residence, education, occupation, wealth); Model B: Multi-level logistic regression model (Covariates included: age, sex, education, occupation, wealth); Model C: Multi-level logistic regression model (Covariates included: age, sex, education, occupation, wealth, area of residence). Reference group: No tobacco use.

Figure 1. Area under the receiver operating characteristic (AU-ROC) curve for tobacco use plotted separately for single and multilevel logistic regression models


Sensitivity
0.00
0.25
0.50
0.75
1.00
0.00
0.25
0.50
0.75
1.00
1-Specificity
model_A ROC area: 0.79
model_B ROC area: 0.8599
model_C ROC area: 0.86
Reference

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2-3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	5
Methods			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	6
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6
Bias	9	Describe any efforts to address potential sources of bias	6
Study size	10	Explain how the study size was arrived at	6
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	6-9
		(*b*) Describe any methods used to examine subgroups and interactions	6-9
		(*c*) Explain how missing data were addressed	6-9
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	6-9
		(*e*) Describe any sensitivity analyses	

Continued on next page

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	9
		(b) Give reasons for non-participation at each stage	
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	
		(b) Indicate number of participants with missing data for each variable of interest	
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	9-10
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	
		(*b*) Report category boundaries when continuous variables were categorized	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	
Discussion			
Key results	18	Summarise key results with reference to study objectives	10
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	10-11
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	11-12
Generalisability	21	Discuss the generalisability (external validity) of the study results	12
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	17

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Geographic variation in tobacco use in India: A population based multi-level study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-033178.R1
Article Type:	Original research
Date Submitted by the Author:	27-Mar-2020
Complete List of Authors:	Singh, Ankur; University of Melbourne, Centre for Health Equity, Melbourne School of Population & Global Health Arora, Monika; Public Health Foundation of India, Health Promotion and Tobacco Control Bentley, Rebecca; University of Melbourne, Centre for Health Equity, Melbourne School of Population & Global Health; University of Melbourne, Centre for Epidemiology and Biostatistics, Melbourne School of Population & Global Health Spittal, Matthew; The University of Melbourne, Melbourne School of Population and Global Health Do, Loc; The University of Adelaide, Australian Research Centre for Population Oral Health (ARCPOH), Adelaide Dental School Grills, Nathan; The University of Melbourne, Nossal Institute for Global Health English, Dallas; University of Melbourne, Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Smoking and tobacco
Keywords:	Tobacco, India, EPIDEMIOLOGY, PUBLIC HEALTH, SOCIAL MEDICINE

Geographic variation in tobacco use in India: A population based multi-level study

List of authors: Ankur Singh (Ph.D.),[1] Monika Arora (Ph.D.),[2] Rebecca Bentley (Ph.D.),[1,3] Matthew J Spittal (Ph.D.),[4] Loc G Do (Ph.D.),[5] Nathan Grills (Ph.D.),[6] Dallas R English (Ph.D)[3]

1. Centre for Health Equity, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Victoria, Australia
2. Health Promotion and Tobacco Control, Public Health Foundation of India, Gurugram, Haryana, India.
3. Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Victoria, Australia
4. Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Victoria, Australia
5. Australian Research Centre for Population Oral Health (ARCPOH), Adelaide Dental School, The University of Adelaide, Adelaide, South Australia, Australia
6. Nossal Institute for Global Health, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Victoria, Australia

Corresponding author:

Dr. Ankur Singh

Centre for Health Equity, Melbourne School of Population and Global Health,

The University of Melbourne,

Level 3, 207 Bouverie Street,

Melbourne, Victoria 3010, Australia

Tel: +61 3 8344 9256 Email: ankur.singh@unimelb.edu.au

Journal: BMJ Open

Publication category: Original Research

Running head: Area-level variations in tobacco use in India

Word count (excluding abstract, references, tables, and figures): 3118 words

Number of references: 43

Conflict of interest: None Declared

Abstract

Objective: This study aims to quantify the extent to which people's use of tobacco products varies by local areas (city-ward/village) across India and the variation in this clustering by tobacco product.

Design: Cross-sectional study

Setting and participants: Data on 73,954 adults across 2,547 city wards and villages was available for analysis from 30 states and two union territories in India.

Methods: We fitted multilevel multivariable logistic regression models to data on adults from the Indian Global Adult Tobacco Survey 2016-2017. We estimated the area-level share of variance in tobacco use (by cigarette smoking, bidi smoking, smokeless tobacco use and dual use [smoking and smokeless tobacco use]) as median odds ratio (MOR) and intra-class coefficients (ICC) adjusting for age, sex, individual wealth, educational attainment, employment status, area of residence and states.

Primary and secondary outcome measures: We included as primary outcomes self-reported any tobacco use, current cigarette smoking, current bidi smoking, current smokeless tobacco use and a derived variable for dual use describing respondents who engaged in both smoking and smokeless tobacco use.

Results: The median risk of an individual using tobacco was 1.64 times greater if a person hypothetically moved from an area of low to high risk of tobacco use (95% CI 1.60 – 1.69). Area-level partitioning of variation differed by tobacco product used. MORs ranged from 1.77 for smokeless tobacco use to 2.53 for dual use.

Conclusion: Tobacco use is highly clustered geographically in India. To be effective in India, policy interventions should be directed at the influence of specific local contextual factors on adult tobacco use. Where people live in India influences their use of tobacco, and

this association may be greater than has been observed in other settings. Tailoring tobacco control policies for local areas in India may, therefore, provide substantial public health benefits.

Strengths and limitations of this study:

- This is the first study from LMICs that has studied variation in tobacco use at local area level using a nationwide representative data.
- By using different measures (ICC, MOR and AUC) we not only inform the extent of variation but we comprehensively examine the degree of clustering, the heterogeneity in outcomes among areas as well as the ability of local areas to classify individuals according to tobacco use.
- We did not incorporate individual-level policy and economic variables related to tobacco use available in the GATS 2016-17 in our analysis. The variables used were non-ecologic and could potentially lead to falsely attributing individual-level measures to areas.

Introduction

Four fifths of the world's current smokers reside in low- and middle-income countries (LMICs), creating enormous societal and public health challenges. [1] The number of deaths from tobacco-related causes and loss of productivity is rapidly increasing in these, often resource poor, settings. [2 3]

The latest Indian Global Adult Tobacco Survey found that nearly 30% of all Indian adults use tobacco. [4] Additionally, the widespread use of smokeless tobacco presents a complex challenge for health systems and tobacco control because of its strong relationship with oral cancerous and pre-cancerous lesions. [5] Despite a nation-wide smokeless tobacco ban implemented in 2013-14, 20% of all tobacco users are smokeless tobacco users. [4] Added to this, the burden of tobacco use in India is disproportionally high among people who are socially disadvantaged. [6-8]

There is consistent evidence that local social and policy contexts shape patterns of tobacco use. [9] Multilevel studies (that simultaneously examine individual- and group-level determinants of health) from The Netherlands, Australia, South Africa, Mexico, Scotland, India, the USA and the UK suggest evidence of an association between area-level contextual factor (such as social disadvantage and local policy environments) and smoking.[10-22] For example, a study of Indian high school students from Mumbai reported the density of tobacco vendors around schools was associated with increased tobacco use by students.[23]

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]

To redress this important gap in evidence, this study aims to quantify the extent to which people's use of tobacco products varies by local areas (city-ward/village) across India and the variation in this clustering by tobacco product.

Methods

Study population

Data on tobacco use in India was obtained from the Global Adult Tobacco Survey (GATS 2) conducted in 2016 and 2017. GATS 2 is a multi-country household tobacco prevalence survey designed to support implementation of tobacco control within study countries.[4] Participants eligible for the survey were non-institutionalised individuals aged 15 years and older. The survey applied a multistage sampling procedure with different sampling hierarchies for urban and rural areas. For urban areas, city wards were the primary sampling unit from which census enumeration blocks, and then households, were selected. In rural areas, the primary sampling units were villages, from which households were selected. A total of 74,037 adults across 2,547 city wards and villages were available for analysis from 30 states and union territories in India. The response rate was 93%. [4]

Data collection

GATS-2 collected data using household and individual questionnaires developed in English and translated into 19 regional languages. The interviewer-administered questionnaires collected data on demographic characteristics, tobacco smoking, smokeless tobacco use, second hand smoke, socioeconomic position, media and knowledge, attitude and perceptions related to tobacco use. More details on sampling procedures and methods of data collection are published elsewhere. [4 30 31]

Outcomes

We included as primary outcomes self-reported any tobacco use, current cigarette smoking, current bidi smoking, current smokeless tobacco use and a derived variable for dual use describing respondents who engaged in both smoking and smokeless tobacco use. Participants were asked 'On average, how many of the following products do you currently smoke each day?. [4 30 31] We categorized those who reported smoking one or more manufactured/rolled tobacco in paper/leaf as current cigarette smokers. Similarly, we identified those who reported smoking one or more bidi as current bidi smokers. Regarding smokeless tobacco use, participants were asked 'Do you currently use smokeless tobacco on a daily basis, less than daily, or not at all? '. [4 30 31] We recorded those answering 'daily' or 'less than daily' as yes for current smokeless tobacco use. Those identified to be both current smokers (cigarette or bidi) and current smokeless tobacco users were identified as dual users. Therefore, we created five binary variables including any tobacco use, current cigarette smokers, current bidi smokers, current smokeless tobacco users and dual users.

Geographic level of aggregation (local areas)

Individuals from urban areas were clustered within city-wards and those in rural areas were clustered within villages. In urban areas, city wards are the units for local government

operations in India, responsible for essential community services including healthcare, education, housing, transport and so on. [32] In rural areas, villages make up the boundary for local panchayat (traditional local self-governance). [32]

Covariates

To account for compositional differences in populations within area-level clusters, we included individual-level demographic characteristics: age (as a continuous variable), sex and socioeconomic position (education: no formal education/less than primary/primary/ secondary or more; occupation: unemployed/labourer/housewife, retired, student/ self-employed/ private/ government; and household-level wealth: quintiles, 1 = lowest, 5 = highest) as covariates in the multilevel regression models. These variables were selected based on a previous study. [7]

Statistical Analysis

We performed the statistical analyses using Stata 15.0 (Statacorp, College Station, TX, USA). We used survey commands to account for the complex survey design and to perform the weighted descriptive analysis. We plotted the prevalence and 95% confidence intervals for any tobacco use and for different types of tobacco use to visually examine their variation by local areas. We fitted multilevel logistic regression models with random intercepts for local areas and fixed slopes with individuals nested in city wards or villages respectively. Multilevel models operationalise studying population-level variations in health outcomes by examining the extent of clustering in health outcomes that exists at the group or contextual level. [24 33-37] Using intra-class correlation coefficients (ICC) and median odds ratios (MOR), we decomposed the variance in tobacco use at city-ward or village levels. The ICC is expressed as a percentage and is interpreted in these analyses as the share of inter-individual variation in health outcome that exists at the group level. For example, an ICC of 8% at the village level means that of all the individual-level variation in tobacco use among rural areas, 8% is attributed to the village level. The higher the individual correlation in health outcomes within a context, the more relevant is the context for understanding individual differences in the health outcome. [25] We estimated the MOR, which describes the area-level variance as an odds ratio, as the median value of the distribution of odds ratios obtained when two individuals with the same covariate values are picked from two different areas, comparing the one from the higher prevalence area to the one from the area with lower prevalence. [24 36 38] In the absence of any area-level variation, the MOR is equal to one. We estimated both MORs and ICCs for binary outcomes as the partition of variance between different levels does not have the same intuitive interpretation as a linear model. [24 38] We estimated ICCs and MORs from intercept only models to examine presence of clustering, and heterogeneity between areas, in the outcomes of tobacco use.

We also applied an alternate method to examine the relevance of area-level contexts for tobacco use by comparing discriminatory accuracies obtained from fitted single-level and multi-level logistic regression models. The area under the receiver operating characteristic curve (AU-ROC) was constructed by plotting the true positive fraction (TPF, sensitivity) against the false positive fraction (FPF, 1 – specificity). It measures the ability of the model to classify individuals with and without the outcome and takes a value between 0.5 and 1.0 where 1.0 is perfect discrimination and 0.5 where covariates have no predictive power. [24]

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 assessed goodness of fit by estimating the changes in the Deviance Information Criterion (DIC). All models were fitted separately for each type of tobacco use (cigarette smoking, bidi smoking, smokeless tobacco use and dual use) to determine any differences in variations in tobacco use according to different types of tobacco use. We performed a sensitivity analysis to examine clustering in tobacco use in city/wards and villages within states by fitting three-level hierarchical models: individual nested within city/wards and villages nested within states.

Patient and Public Involvement

No patients or public were involved in this study.

Results

We analysed data for 73,954 individuals (99.9%) of the 74,037 survey participants. We did not analyse data on 83 participants due to missing covariates data. Table 1 shows descriptive characteristics of the sample according to residence status. 28% of adults used tobacco products. The prevalence of smokeless tobacco use was 18.6% (Table 1). Plots for prevalence and 95% CI for any tobacco use and different types of tobacco use by local areas showed substantial variations (Supplementary appendix).

Intercept only models (null models with no covariate adjustment) estimated 22% (95% CI: 20, 24) of any tobacco use was clustered at the city-ward/village level. Cigarette smoking was clustered 31%, bidi smoking at 28%, dual use at 40% and smokeless tobacco at 36% respectively (estimates not reported in the tables). For each outcome, the AUC increased when multilevel logistic regression models were fitted. The AUC increased to 0.86 in Model B as compared to 0.79 in a 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).

After including all individual-level covariates, the proportion of variance attributable to the areas remained at 21% (95% CI: 20, 22) (Table 2). Correspondingly, the median odds ratio for was 2.43 (95% CI: 2.35, 2.52). 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. Including area of residence and state in Model C substantially reduced the estimates of proportion of variance attributable to areas and the respective median odds ratios. The proportion of variance for any tobacco use reduced from 21% to 7.6% and corresponding median ratio from 2.42 to 1.64. Sensitivity analysis confirmed our findings of high clustering in any tobacco use within city-ward or villages from the same state (Supplementary Appendix pp.21).

The decrease in DIC values between the single-level models and multilevel models including covariates suggested better model fit (Table 2).

Among the different types of tobacco use, the highest ICC (22%; 95% CI: 19, 26) and MOR (2.53; 95% CI: 2.32, 2.74) were for dual use and the lowest for SLT use ((ICC: 10%; 95% CI: 9, 11), (MOR: 1.77; 95% CI: 1.71, 1.83)) (Table 3). Similar to any tobacco use, substantial reductions in estimates of ICC and MOR were observed upon inclusion of state and area of residence in Model C compared with Model B.

Discussion

We found substantial variation in tobacco use across local areas in India. Individual-level social and demographic characteristics were not able to explain the high area-level variations in tobacco. Including states and area of residence explained substantial area-level variation in tobacco use. However, the remaining variation in tobacco use was still high, indicating the importance of local areas. The degree of area-level variation in tobacco use differed according to the types of tobacco product. Dual use (smoking and smokeless) had the highest geographic clustering.

Strengths and Limitations

This study had several strengths and limitations. To the best of our knowledge this is the first study from LMICs that has studied variation in tobacco use at local area level using a nationwide representative data. [4] By using different measures (ICC, MOR and AUC) we not only inform the extent of variation but we comprehensively examine the degree of clustering, the heterogeneity in outcomes among areas as well as the ability of local areas to classify individuals according to tobacco use. [24 33-36] This study also has limitations. 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 the respondent's perceptions rather than objective measures of availability and implementation of policies in local areas and because this data was only gathered from smokers. The non-ecologic nature of these variables could lead to falsely attributing individual-level measures to area levels (the atomistic fallacy). [37]

Discussion in context of current evidence

Our findings of high variations in tobacco use among local areas is new. A multilevel study on societal determinants of tobacco use from Scotland found no evidence of clustering in tobacco use at the area level. [18] Other multilevel studies have not presented measures of variance, which limits comparisons.[10-22] 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.

Tobacco specific variations in the values of ICC and MOR highlight potential differences in the relevance of contexts by type of tobacco product used. Evidence from other studies suggests that while wealthier and more educated individuals have higher odds of cigarette smoking than their disadvantaged counterparts, disadvantaged individuals have higher odds of bidi smoking and smokeless tobacco use. [6 7] Our study highlights the presence of both individual and geographic socioeconomic inequalities in tobacco use by product. For example, we observed a higher effect of individual social and demographic characteristics in smokeless tobacco use when compared to cigarette smoking and bidi smoking for contextual effects (change in ICC from 36% in null model to 31% in adjusted model). In addition, the proportion of variation for all types of tobacco use was markedly explained by adding states into the model. This emphasizes the role of cultural and regional diversity within India in determining tobacco use.[39] Both ICCs from the three-level hierarchical models and odds ratios estimated from regression models confirmed pivotal role played by states in geographic inequities in tobacco use in India (see Supplementary Appendix).

Research and policy implications

Given the role of contexts in shaping individual health behaviours, this study builds a framework for operationalizing a contextual thinking in tobacco control activities, particularly in LMICs where social norms and cultural aspects may differ from high-income countries. High general contextual effects of local areas for tobacco use necessitates a thorough examination of factors at the area-level that may be causally associated with individual tobacco use as well as those which can explain the high variations in tobacco use among local areas. This may only be possible if either data on individual-level tobacco use is linked with small area characteristics, or if future population-based surveys collect both area- and individual-level data relevant to tobacco use. 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. Furthermore, current findings build the platform for more robust population-based studies that collectively examine area- and individual-level determinants of tobacco use in India and other LMICs.

This study has several policy implications. Our findings confirm that context plays an important role in determining use of tobacco. India's Cigarettes and other Tobacco Products Act (COTPA) is a national law, which is in line with World Health Organization's Framework Convention on Tobacco Control. States at sub-national level are responsible for implementing various tobacco control policy measures under COTPA. Comparison of GATS-2 and GATS-1, and household surveys, has highlighted changes in prevalence of tobacco use due to differential implementation of these measures.[40 41] States are also allowed to develop context specific information, education and communication resources to match the local needs. [41 42] Therefore, health promotion and tobacco control interventions must be designed for contexts and applied contextually rather than being individually oriented.[9 43] There is the potential to enhance National Tobacco Control Program's (NTCP) implementation at city-ward, village and block level as well.[41] 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. 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 disadvantange to reduce geographic inequalities.

Conclusion

Where people live in India influences their use of tobacco, and this association may be greater than has been observed in other settings. Tailoring tobacco control policies for local areas in India may, therefore, provide substantial public health benefits.

References

1. WHO. Tobacco: Fact sheet Geneva: World Health Organization; 2018 [Available from: https://www.who.int/en/news-room/fact-sheets/detail/tobacco accessed 10 December 2018.
2. Hipple Walters B, Petrea I, Lando H. Tobacco Control in Low- and Middle-Income Countries: Changing the Present to Help the Future. *Journal of Smoking Cessation* 2018:1-2. doi: 10.1017/jsc.2018.4 [published Online First: 03/14]
3. Gilmore AB, Fooks G, Drope J, et al. Exposing and addressing tobacco industry conduct in low-income and middle-income countries. *The Lancet* 2015;385(9972):1029-43. doi: 10.1016/S0140-6736(15)60312-9
4. Tata Institute of Social Sciences (TISS) MaMoHaFW, Government of India. Global Adult Tobacco Survey GATS 2 India 2016-17. New Delhi: Tata Institute of Social Sciences (TISS), Mumbai and Ministry of Health and Family Welfare, Government of India, 2018.
5. Gupta PC, Arora M, Sinha D, et al. Smokeless tobacco and public health in India. *Ministry of Health & Family Welfare, Government of India; New Delhi* 2016
6. Corsi DJ, Subramanian SV. Divergent socio-economic gradients in smoking by type of tobacco use in India. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease* 2014;18(1):122-4. doi: 10.5588/ijtld.13.0246 [published Online First: 2013/12/25]
7. Singh A, Arora M, English DR, et al. Socioeconomic Gradients in Different Types of Tobacco Use in India: Evidence from Global Adult Tobacco Survey 2009-10. *Biomed Res Int* 2015;2015:837804. doi: 10.1155/2015/837804 [published Online First: 2015/08/15]
8. Bhan N, Srivastava S, Agrawal S, et al. Are socioeconomic disparities in tobacco consumption increasing in India? A repeated cross-sectional multilevel analysis. *BMJ open* 2012;2(5) doi: 10.1136/bmjopen-2012-001348 [published Online First: 2012/10/02]
9. Link BG, Phelan J. Social Conditions as Fundamental Causes of Disease. *Journal of Health and Social Behavior* 1995;35:80-94. doi: Doi 10.2307/2626958
10. Rachele JN, Wood L, Nathan A, et al. Neighbourhood disadvantage and smoking: Examining the role of neighbourhood-level psychosocial characteristics. *Health Place* 2016;40:98-105. doi: 10.1016/j.healthplace.2016.04.012 [published Online First: 2016/05/27]
11. Ellaway A, Macintyre S. Are perceived neighbourhood problems associated with the likelihood of smoking? *Journal of epidemiology and community health* 2009;63(1):78-80. doi: 10.1136/jech.2007.068767 [published Online First: 2008/12/18]
12. Morris T, Manley D, Van Ham M. Context or composition: How does neighbourhood deprivation impact upon adolescent smoking behaviour? *PloS one* 2018;13(2):e0192566. doi: 10.1371/journal.pone.0192566 [published Online First: 2018/02/09]
13. Mayne SL, Auchincloss AH, Moore KA, et al. Cross-sectional and longitudinal associations of neighbourhood social environment and smoking behaviour: the multiethnic study of atherosclerosis. *Journal of epidemiology and community health* 2017;71(4):396-403. doi: 10.1136/jech-2016-207990 [published Online First: 2016/11/26]
14. Chuang YC, Cubbin C, Ahn D, et al. Effects of neighbourhood socioeconomic status and convenience store concentration on individual level smoking. *Journal of*

epidemiology and community health 2005;59(7):568-73. doi: 10.1136/jech.2004.029041 [published Online First: 2005/06/21]
15. Fleischer NL, Lozano P, Arillo Santillan E, et al. The impact of neighbourhood violence and social cohesion on smoking behaviours among a cohort of smokers in Mexico. *Journal of epidemiology and community health* 2015;69(11):1083-90. doi: 10.1136/jech-2014-205115 [published Online First: 2015/06/06]
16. Turrell G, Hewitt BA, Miller SA. The influence of neighbourhood disadvantage on smoking cessation and its contribution to inequalities in smoking status. *Drug Alcohol Rev* 2012;31(5):645-52. doi: 10.1111/j.1465-3362.2012.00452.x [published Online First: 2012/04/18]
17. van Lenthe FJ, Mackenbach JP. Neighbourhood and individual socioeconomic inequalities in smoking: the role of physical neighbourhood stressors. *Journal of epidemiology and community health* 2006;60(8):699-705. doi: 10.1136/jech.2005.043851 [published Online First: 2006/07/15]
18. Shareck M, Ellaway A. Neighbourhood crime and smoking: the role of objective and perceived crime measures. *BMC public health* 2011;11:930. doi: 10.1186/1471-2458-11-930 [published Online First: 2011/12/16]
19. Fleischer NL, Thrasher JF, Saenz de Miera Juarez B, et al. Neighbourhood deprivation and smoking and quit behaviour among smokers in Mexico: findings from the ITC Mexico Survey. *Tobacco control* 2015;24 Suppl 3:iii56-iii63. doi: 10.1136/tobaccocontrol-2013-051495 [published Online First: 2014/08/30]
20. Lau YK, Tam J, Fleischer NL, et al. Neighbourhood deprivation, smoking, and race in South Africa: A cross-sectional analysis. *Prev Med Rep* 2018;11:202-08. doi: 10.1016/j.pmedr.2018.07.001 [published Online First: 2018/07/13]
21. Timmermans EJ, Veldhuizen EM, Snijder MB, et al. Neighbourhood safety and smoking in population subgroups: The HELIUS study. *Preventive medicine* 2018;112:111-18. doi: 10.1016/j.ypmed.2018.04.012 [published Online First: 2018/04/15]
22. Reijneveld SA. Neighbourhood socioeconomic context and self reported health and smoking: a secondary analysis of data on seven cities. *Journal of epidemiology and community health* 2002;56(12):935-42. [published Online First: 2002/12/04]
23. Mistry R, Pednekar M, Pimple S, et al. Banning tobacco sales and advertisements near educational institutions may reduce students' tobacco use risk: evidence from Mumbai, India. *Tobacco control* 2015;24(e1):e100-7. doi: 10.1136/tobaccocontrol-2012-050819 [published Online First: 2013/08/21]
24. Merlo J, Wagner P, Ghith N, et al. An Original Stepwise Multilevel Logistic Regression Analysis of Discriminatory Accuracy: The Case of Neighbourhoods and Health. *PloS one* 2016;11(4):e0153778. doi: 10.1371/journal.pone.0153778 [published Online First: 2016/04/28]
25. Merlo J, Wagner P, Austin PC, et al. General and specific contextual effects in multilevel regression analyses and their paradoxical relationship: A conceptual tutorial. *SSM - Population Health* 2018;5:33-37. doi: 10.1016/j.ssmph.2018.05.006
26. Diez Roux AV. The study of group-level factors in epidemiology: rethinking variables, study designs, and analytical approaches. *Epidemiologic reviews* 2004;26(1):104-11. doi: 10.1093/epirev/mxh006 [published Online First: 2004/07/06]
27. Merlo J. Multilevel analytical approaches in social epidemiology: measures of health variation compared with traditional measures of association. *Journal of epidemiology and community health* 2003;57(8):550-2. [published Online First: 2003/07/29]

28. Merlo J, Chaix B, Yang M, et al. A brief conceptual tutorial on multilevel analysis in social epidemiology: interpreting neighbourhood differences and the effect of neighbourhood characteristics on individual health. *Journal of epidemiology and community health* 2005;59(12):1022-8. doi: 10.1136/jech.2004.028035 [published Online First: 2005/11/16]
29. Merlo J, Viciana-Fernandez FJ, Ramiro-Farinas D, et al. Bringing the individual back to small-area variation studies: a multilevel analysis of all-cause mortality in Andalusia, Spain. *Soc Sci Med* 2012;75(8):1477-87. doi: 10.1016/j.socscimed.2012.06.004 [published Online First: 2012/07/17]
30. Palipudi KM, Morton J, Hsia J, et al. Methodology of the Global Adult Tobacco Survey - 2008-2010. *Global health promotion* 2016;23(2 Suppl):3-23. doi: 10.1177/1757975913499800 [published Online First: 2013/09/18]
31. GATS. Global Adult Tobacco Survey (GATS) India Report 2009-2010. : Ministry of Health and Family Welfare, India, 2010.
32. Office of the Registrar General & Census Commissioner I. Administrative Divisions New Delhi, India: Ministry of Home Affairs, Government of India; 2018 [Available from: http://censusindia.gov.in/Census_And_You/Administrative_division.aspx accessed 30/11/18 2018.
33. Subramanian SV. The relevance of multilevel statistical methods for identifying causal neighborhood effects. *Soc Sci Med* 2004;58(10):1961-7. doi: 10.1016/S0277-9536(03)00415-5 [published Online First: 2004/03/17]
34. Subramanian SV, Jones K, Kaddour A, et al. Revisiting Robinson: the perils of individualistic and ecologic fallacy. *International journal of epidemiology* 2009;38(2):342-60; author reply 70-3. doi: 10.1093/ije/dyn359 [published Online First: 2009/01/31]
35. Merlo J, Chaix B, Yang M, et al. A brief conceptual tutorial of multilevel analysis in social epidemiology: linking the statistical concept of clustering to the idea of contextual phenomenon. *Journal of epidemiology and community health* 2005;59(6):443-9. doi: 10.1136/jech.2004.023473 [published Online First: 2005/05/25]
36. Merlo J, Ohlsson H, Lynch KF, et al. Individual and collective bodies: using measures of variance and association in contextual epidemiology. *Journal of epidemiology and community health* 2009;63(12):1043-8. doi: 10.1136/jech.2009.088310 [published Online First: 2009/08/12]
37. Singh A, Harford J, Peres MA. Investigating societal determinants of oral health-Opportunities and challenges in multilevel studies. *Community dentistry and oral epidemiology* 2018;46(4):317-27. doi: 10.1111/cdoe.12369 [published Online First: 2018/02/21]
38. Merlo J, Chaix B, Ohlsson H, et al. A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *Journal of epidemiology and community health* 2006;60(4):290-7. doi: 10.1136/jech.2004.029454 [published Online First: 2006/03/16]
39. Sarkar BK, Reddy KS. Priorities for tobacco control research in India. *Addiction* 2012;107(12):2066-68. doi: 10.1111/j.1360-0443.2012.03942.x
40. 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):928-33. doi: 10.15585/mmwr.mm6841a1 [published Online First: 2019/10/18]
41. 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):103-10. doi: 10.1136/tobaccocontrol-2018-054621 [published Online First: 2018/12/17]
42. 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.
43. Phelan JC, Link BG, Tehranifar P. Social conditions as fundamental causes of health inequalities: theory, evidence, and policy implications. *J Health Soc Behav* 2010;51 Suppl(1_suppl):S28-40. doi: 10.1177/0022146510383498 [published Online First: 2010/12/22]

Footnotes

Author's contribution:
AS conceptualised the study, acquisition of data, analysed the data, interpretation of results, led the manuscript preparation and the submission process
MA contributed to interpretation of local policy implications of the results and drafts of the manuscript
RB and NG contributed by critical inputs on multiple draft of the manuscript and interpretation of results
MS and LD contributed by development of analytical framework, interpretation of results and revision of manuscript
DE contributed by reviewing early and advanced drafts of the manuscript, development of analysis plan and interpretation of results.

Funding statement: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. Matthew Spittal is a recipient of an Australian Research Council Future Fellowship (project number FT180100075) funded by the Australian Government
Competing interests: None declared.
Declaration of Interest: Authors have no conflict of interest to declare.
Patient consent for publication: Not required.
Data sharing statement: De-identified data is available from GTSS: https://www.cdc.gov/tobacco/global/gtss/index.htm
Ethics approval: No formal ethical approval was required.
Provenance and peer review: Not commissioned; externally peer reviewed.

Tables and figures

Table 1. Descriptive characteristics of the sample (n=73,954)

Variable	Categories	Percentage
Age (years)	15 to 30	41.7
	31 to 45	29.7
	46 to 60	17.7
	61 to 75	8.9
	76 and above	1.8
Sex	Male	51.1
	Female	48.9
Wealth	Poorer	23.4
	Poor	36.5
	Middle	15.0
	Rich	12.2
	Richer	12.9
Education	No formal education	26.4
	Less than primary	9.2
	Primary	28.2
	Secondary or more	36.2
Occupation	Unemployed	4.3
	Labour	21.2
	Housewife/ Retired/ Student	44.1
	Self	19.4
	Private	8.3
	Government job	2.7
Area of residence	Urban	34.5
	Rural	65.5
Tobacco use	Non- user	72.2
	Cigarette smoking	1.3
	Bidi smoking	4.6
	Smokeless tobacco use	18.6
	Dual Use	3.3

Weighted percentages (Using Survey Weights)

Table 2. Multilevel logistic regression models for any tobacco use among Indian adults (n=73,954 individuals nested in 2547 city wards and villages).

	Model A Estimate 95% CI	Model B Estimate 95% CI	Model C Estimate 95% CI
AUC	0.79	0.86	0.86
AUC change		0.07	0
Variance		0.87	0.27
		(0.80, 0.94)	(0.24, 0.30)
ICC		21%	8%
		(20, 22)	(7, 9)
MOR		2.43	1.64
		(2.35, 2.52)	(1.60, 1.69)
DIC	71171.7	66619.6	64702.3
DIC change		-4552.1	-1917.3

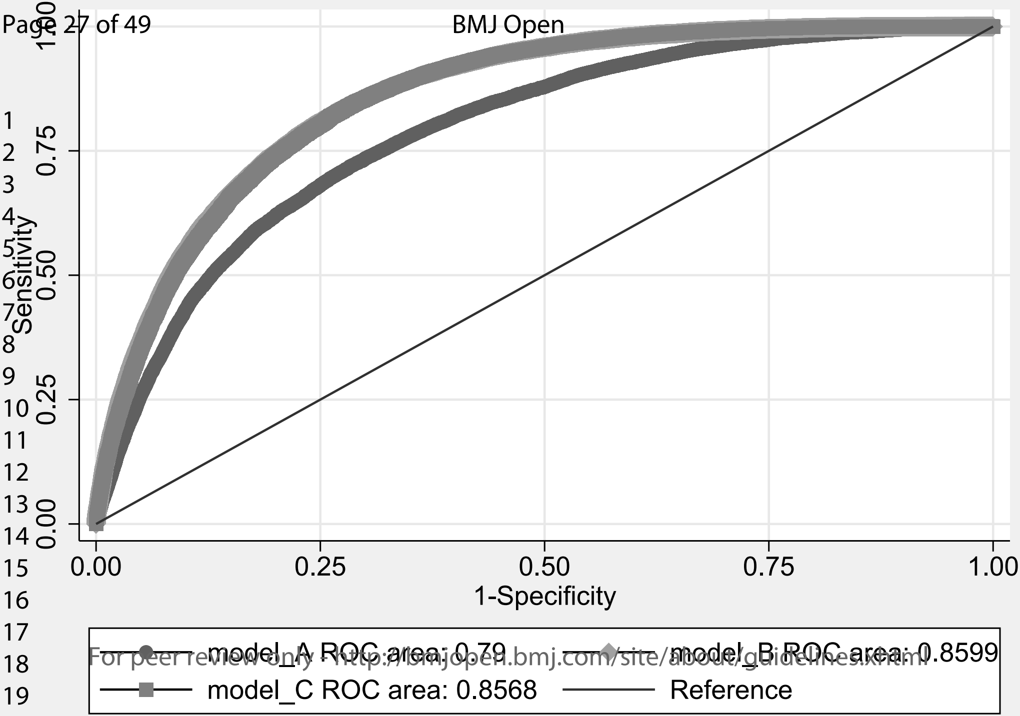
Model A: Single-level logistic regression model (Covariates included: age, sex, area of residence, education, occupation, wealth); Model B: Multi-level logistic regression model (Covariates included: age, sex, education, occupation, wealth); Model C: Multi-level logistic regression model (Covariates included: age, sex, education, occupation, wealth, area of residence and states). Reference group: No tobacco use.

Table 3. Multilevel logistic regression models for different types of tobacco use among Indian adults (n=73,954 individuals nested in 2547 city wards and villages).

	Model A Estimate 95% CI	Model B Estimate 95% CI	Model C Estimate 95% CI
Cigarettes (n=54,648)			
Variance		1.44 (1.19, 1.70)	0.53 (0.38, 0.68)
ICC (%)		30 (27, 34)	14 (11, 18)
MOR		3.14 (2.83, 3.46)	2.00 (2.82, 3.46)
DIC	10630.4	10175.0	9480.5
DIC change		-455.4	-694.5
AUC	0.68	0.70	0.69
AUC change		0.02	-0.01
Bidi (n=56,814)			
Variance		1.53 (1.33, 1.72)	0.65 (0.53, 0.76)
ICC (%)		32 (29, 35)	16 (14,19)
MOR		3.25 (3.01, 3.49)	2.15 (2.01, 2.30)
DIC	18822.5	17680.8	16765.4
DIC change		-1141.7	-915.4
AUC	0.89	0.95	0.94
AUC change		0.06	-0.01
SLT (n=66,089)			
Variance		1.46 (1.34, 1.59)	0.36 (0.31, 0.40)
ICC (%)		31 (29, 33)	10 (9, 11)
MOR		3.17 (3.01, 3.32)	1.77 (1.71, 1.83)
DIC	56207.3	51179.1	48915.1
DIC change		-5028.1	-2264.0
AUC	0.76	0.87	0.86
AUC change		0.11	-0.01
Dual use (n=55,522)			
Variance		2.41 (2.09, 2.72)	0.95 (0.78, 1.12)
ICC (%)		42 (39, 45)	22 (19, 26)
MOR		4.39 (3.96, 4.82)	2.53 (2.32, 2.74)
DIC	14335.7	12989.8	12045.9
DIC change		-1345.9	-943.9
AUC	0.88	0.96	0.95
AUC change		0.08	-0.01

Model A: Single-level logistic regression model (Covariates included: age, sex, area of residence, education, occupation, wealth); Model B: Multi-level logistic regression model (Covariates included: age, sex, education, occupation, wealth); Model C: Multi-level logistic regression model (Covariates included: age, sex, education, occupation, wealth, area of residence and states). Reference group: No tobacco use.

Figure 1. Area under the receiver operating characteristic (AU-ROC) curve for tobacco use plotted separately for single and multilevel logistic regression models

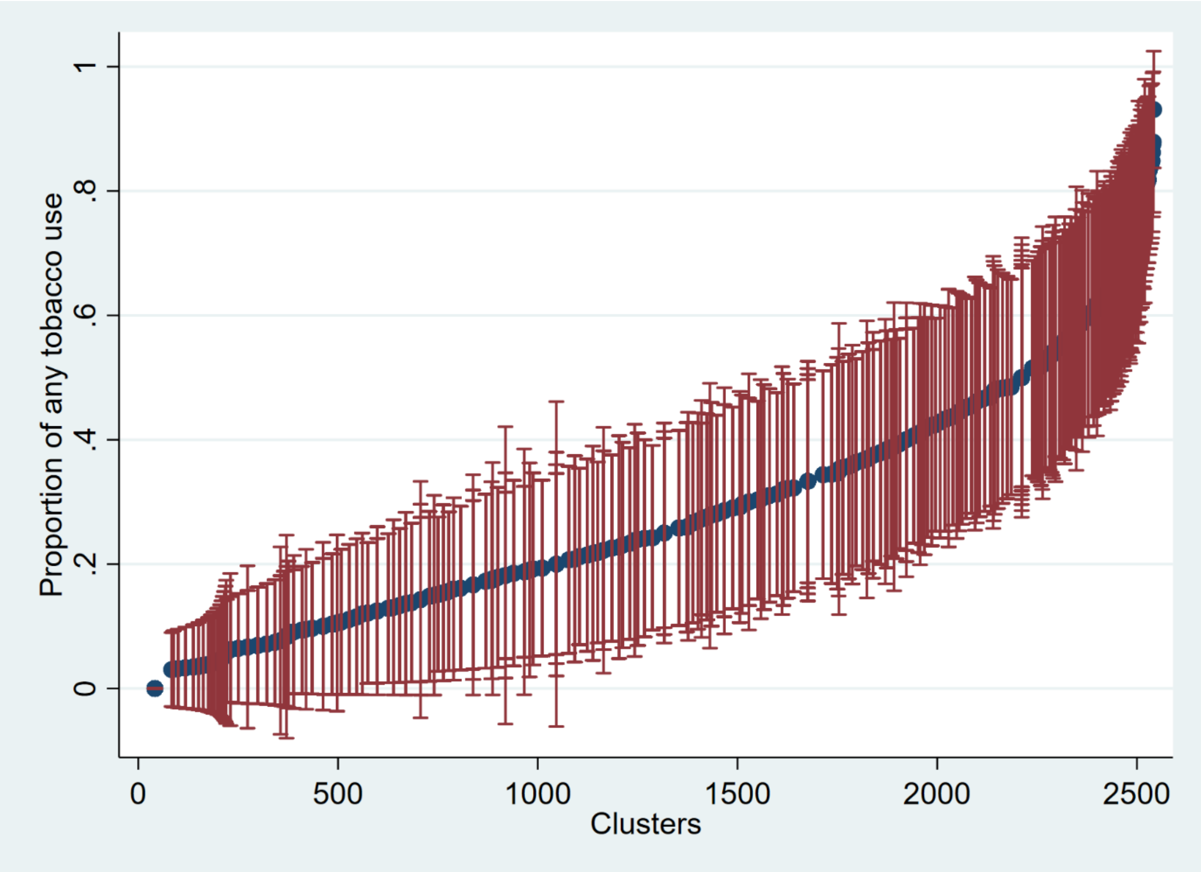

1.00
0.75
0.50
0.25
0.00
Sensitivity
0.00
0.25
0.50
0.75
1.00
1-Specificity
model_A ROC area: 0.79
model_B ROC area: 0.8599
model_C ROC area: 0.8568
Reference

Supplementary Appendix

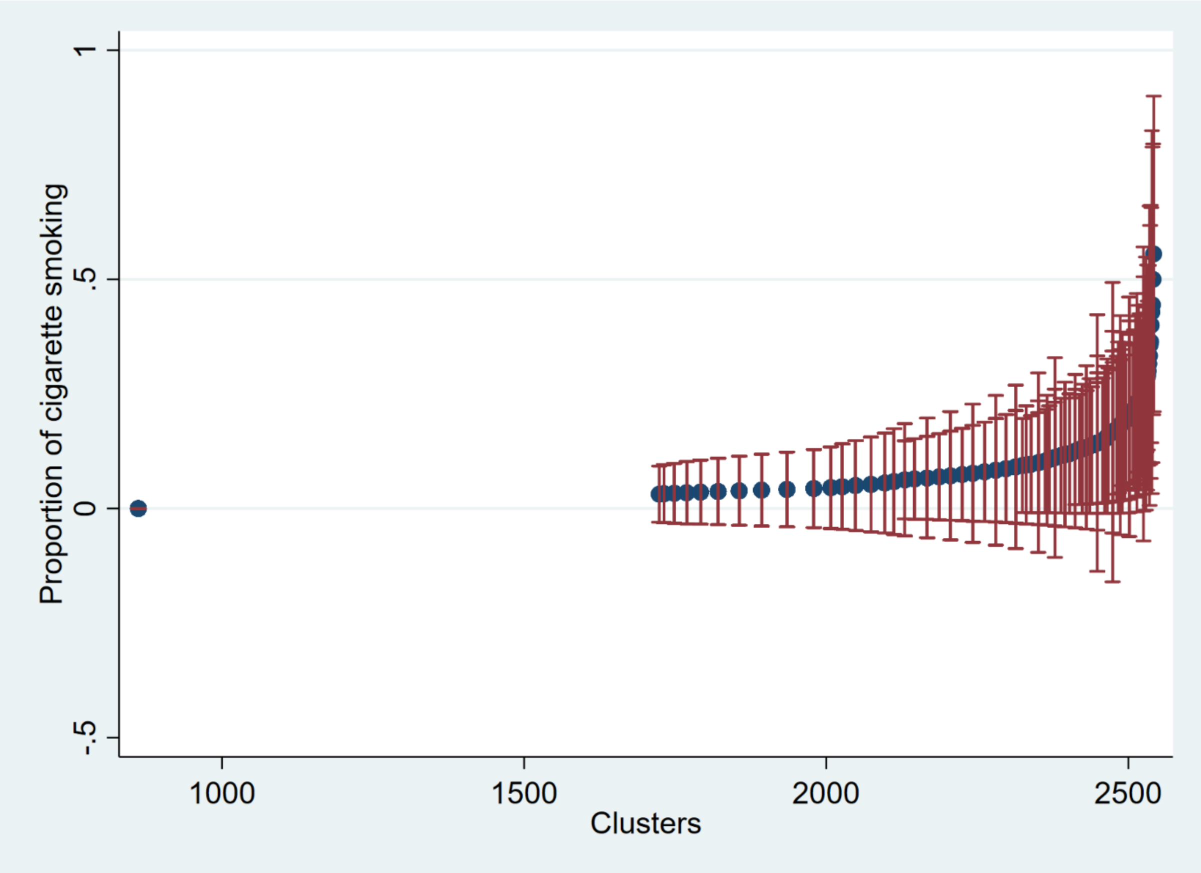
Contents

Prevalence and 95% confidence intervals of any tobacco use by local areas 2
Prevalence and 95% confidence intervals of cigarette smoking by local areas 3
Prevalence and 95% confidence intervals of bidi smoking by local areas 4
Prevalence and 95% confidence intervals of SLT use by local areas 5
Prevalence and 95% confidence intervals of Dual use by local areas 6
Odds ratios for any tobacco use obtained from multilevel multivariable logistic regression models ... 7
Odds ratios for cigarette smoking obtained from multilevel multivariable logistic regression models 9
Odds ratios for bidi smoking obtained from multilevel multivariable logistic regression models 11
Odds ratios for smokeless tobacco use obtained from multilevel multivariable logistic regression models 13
Odds ratios for dual use obtained from multilevel multivariable logistic regression models 15
Area under the receiver operating characteristic (AU-ROC) curve for *cigarette smoking* plotted separately for single and multilevel logistic regression models 17
Area under the receiver operating characteristic (AU-ROC) curve for *bidi smoking* plotted separately for single and multilevel logistic regression models 18
Area under the receiver operating characteristic (AU-ROC) curve for *SLT use* plotted separately for single and multilevel logistic regression models 19
Area under the receiver operating characteristic (AU-ROC) curve for *dual use* plotted separately for single and multilevel logistic regression models 20
Table showing intraclass correlation coefficients obtained from three level hierarchical models with individual nested within city-wards/villages nested within states 21

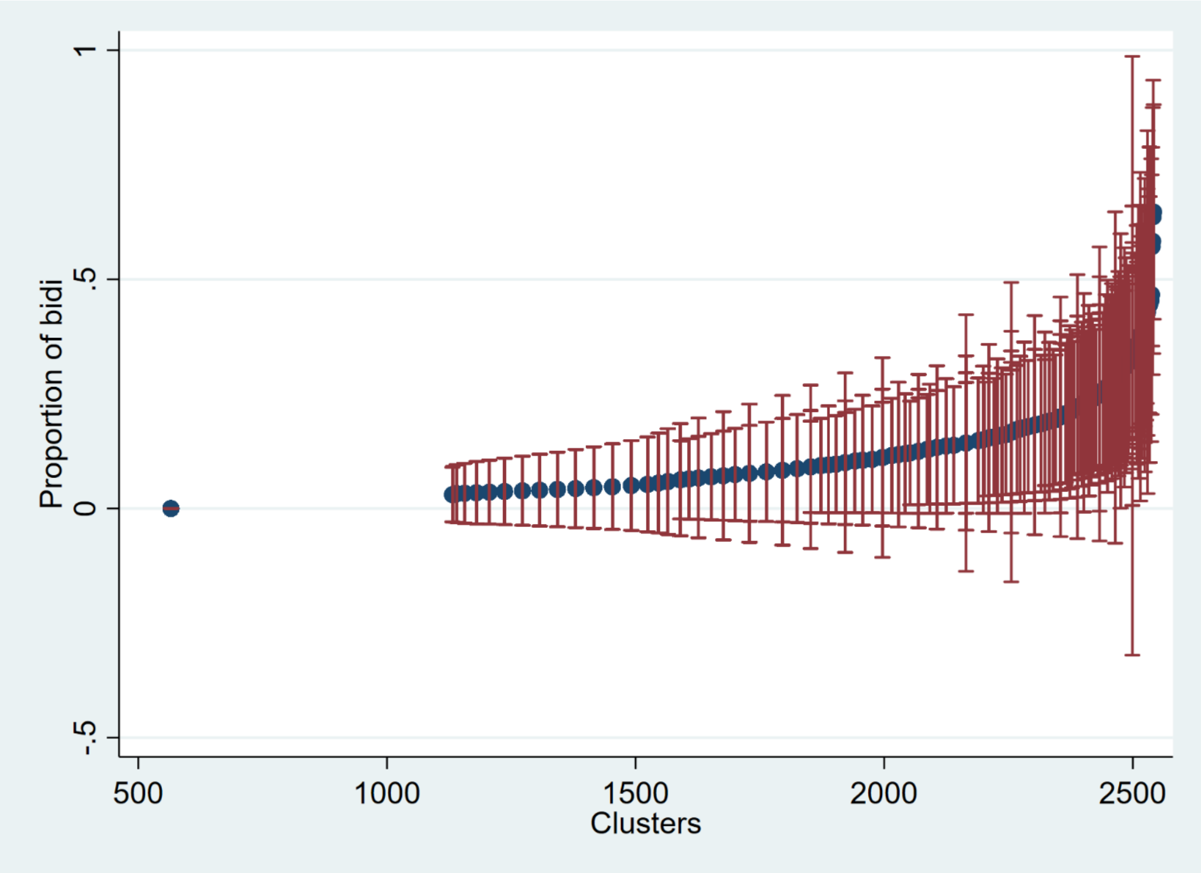
Prevalence and 95% confidence intervals of any tobacco use by local areas



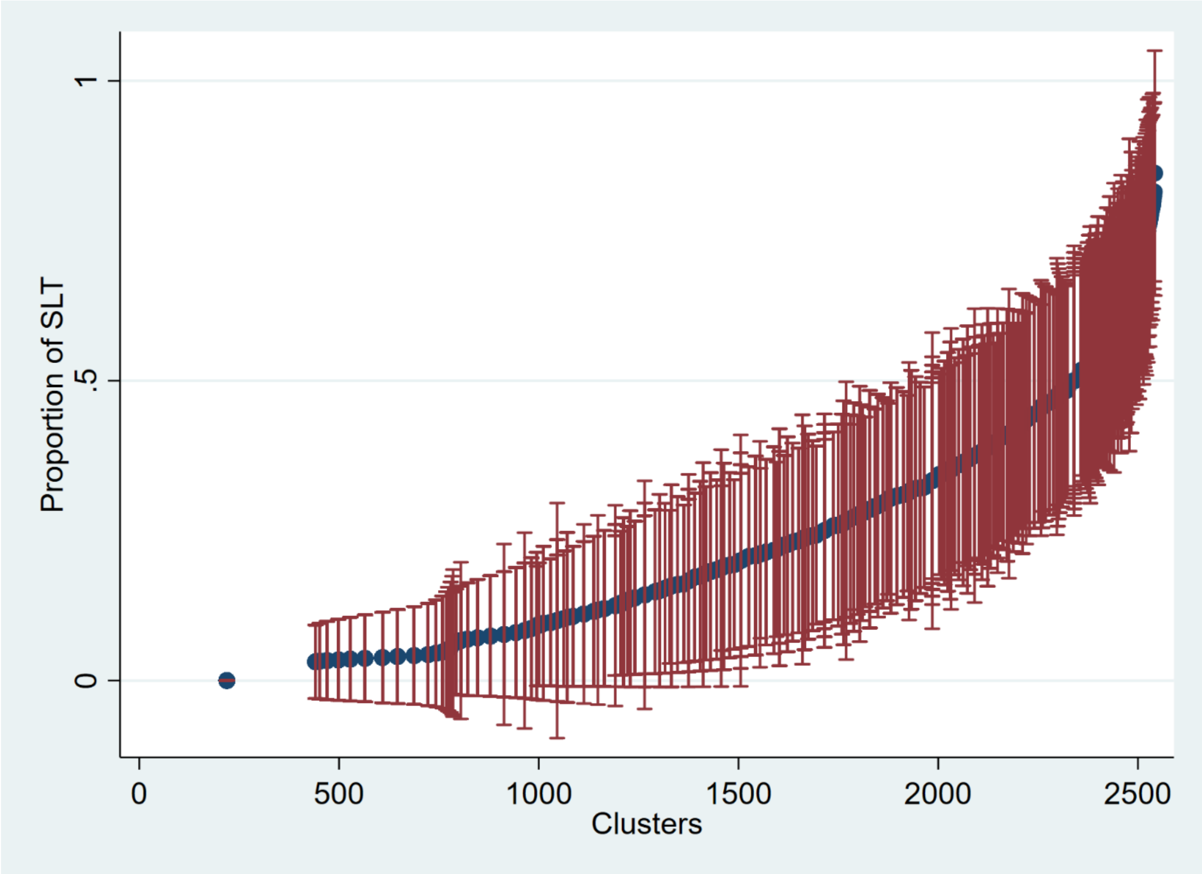
Prevalence and 95% confidence intervals of cigarette smoking by local areas



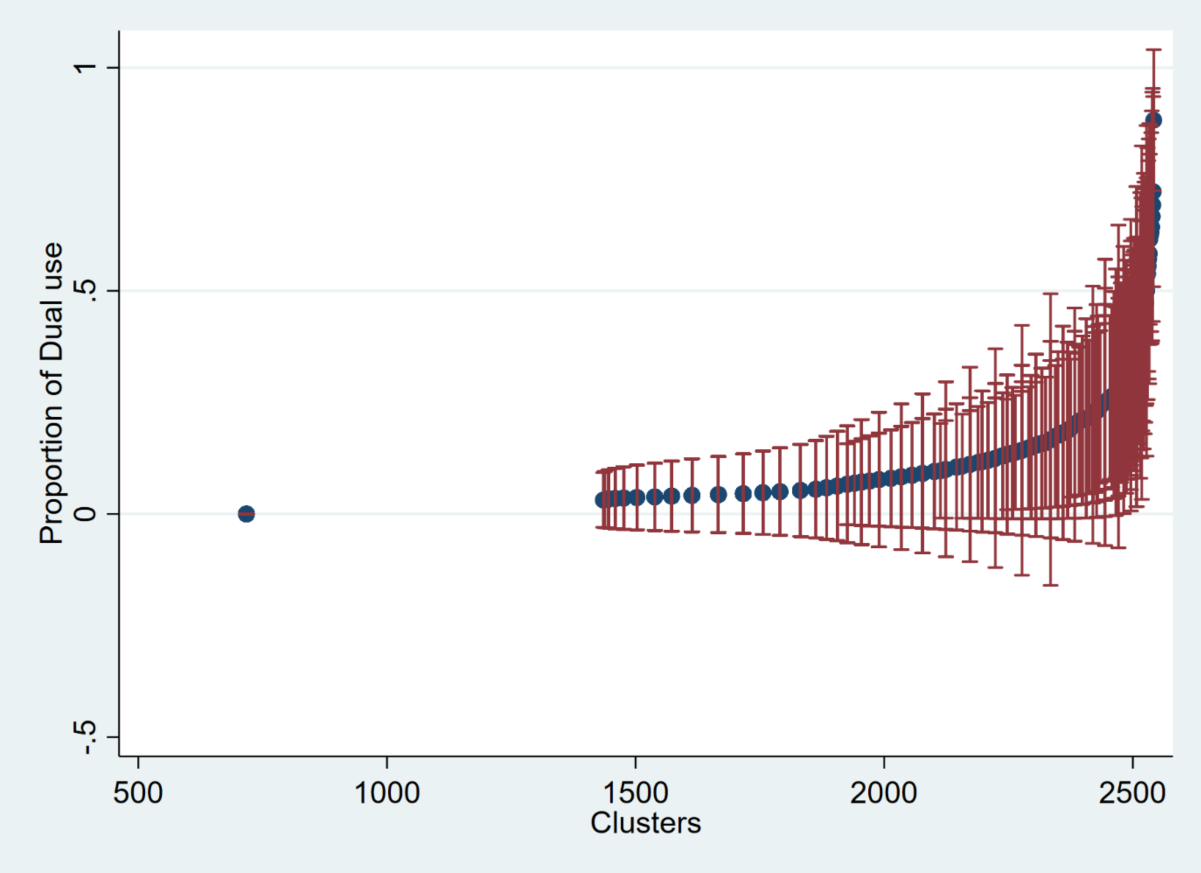
Prevalence and 95% confidence intervals of bidi smoking by local areas



Prevalence and 95% confidence intervals of SLT use by local areas



Prevalence and 95% confidence intervals of Dual use by local areas



Odds ratios for any tobacco use obtained from multilevel multivariable logistic regression models

Covariates	Categories	Odds Ratio	95% CI	
Age		1.02	1.02	1.03
Sex	Male			
	Female	0.23	0.21	0.24
Wealth	Poorer			
	Poor	0.85	0.81	0.90
	Middle	0.74	0.69	0.80
	Rich	0.58	0.53	0.63
	Richer	0.43	0.40	0.48
Education	No formal education			
	Less than primary	0.86	0.80	0.92
	Primary less than secondary	0.68	0.64	0.73
	Secondary and above	0.37	0.35	0.40
Occupation	Unemployed			
	Labourer	1.92	1.74	2.12
	Housewife/ Retired/ Student	0.66	0.60	0.73
	Self	1.54	1.40	1.70
	Private	1.53	1.37	1.72
	Government	1.14	1.00	1.30
Area of residence	Urban			
	Rural	1.11	1.04	1.19
States	Jammu & Kashmir			
	Himachal Pradesh	0.81	0.64	1.03
	Punjab	0.71	0.55	0.90
	Chandigarh	0.73	0.56	0.94
	Uttarakhand	1.44	1.15	1.81
	Haryana	1.20	0.95	1.52
	Delhi	1.18	0.92	1.51
	Rajasthan	1.06	0.85	1.32
	Uttar Pradesh	2.34	1.90	2.88
	Chhattisgarh	3.06	2.43	3.86
	Madhya Pradesh	1.78	1.44	2.21
	West Bengal	1.92	1.55	2.39
	Jharkhand	2.48	1.96	3.13
	Odisha	3.87	3.06	4.90
	Bihar	1.16	0.94	1.45
	Sikkim	0.86	0.66	1.12
	Arunachal Pradesh	3.74	2.90	4.82
	Nagaland	5.06	3.94	6.48
	Manipur	9.46	7.37	12.14
	Mizoram	4.64	3.61	5.95
	Tripura	11.87	9.21	15.29
	Meghalaya	3.06	2.38	3.92
	Assam	4.36	3.52	5.41
	Gujarat	1.26	1.01	1.58
	Maharashtra	1.39	1.12	1.72

	Goa	0.44	0.34	0.58
	Andhra Pradesh	0.50	0.39	0.63
	Telangana	0.68	0.53	0.87
	Karnataka	1.04	0.84	1.30
	Kerala	0.57	0.44	0.73
	Tamil Nadu	0.80	0.64	1.00
	Puducherry	0.53	0.41	0.68

Odds ratios for cigarette smoking obtained from multilevel multivariable logistic regression models

Covariates	Categories	Odds Ratio	95% CI	
Age		1.01	1.00	1.01
Sex	Male	0.03	0.02	0.04
	Female			
Wealth	Poorer			
	Poor	1.31	1.03	1.68
	Middle	1.80	1.38	2.35
	Rich	1.86	1.40	2.46
	Richer	1.80	1.36	2.40
Education	No formal education			
	Less than primary	1.14	0.86	1.51
	Primary less than secondary	1.24	0.98	1.56
	Secondary and above	1.02	0.81	1.30
Occupation	Unemployed			
	Labourer	1.70	1.24	2.34
	Housewife/ Retired/ Student	0.51	0.37	0.72
	Self	1.51	1.12	2.05
	Private	1.54	1.11	2.14
	Government	1.32	0.94	1.87
Area of residence	Urban			
	Rural	0.66	0.56	0.77
States	Jammu & Kashmir			
	Himachal Pradesh	0.13	0.08	0.21
	Punjab	0.08	0.04	0.13
	Chandigarh	0.10	0.06	0.16
	Uttarakhand	0.16	0.10	0.26
	Haryana	0.06	0.03	0.10
	Delhi	0.13	0.09	0.21
	Rajasthan	0.05	0.03	0.09
	Uttar Pradesh	0.05	0.03	0.10
	Chhattisgarh	0.03	0.01	0.09
	Madhya Pradesh	0.04	0.02	0.08
	West Bengal	0.19	0.12	0.30
	Jharkhand	0.21	0.12	0.37
	Odisha	0.09	0.04	0.19
	Bihar	0.02	0.01	0.06
	Sikkim	0.46	0.30	0.70
	Arunachal Pradesh	0.59	0.37	0.96
	Nagaland	0.07	0.03	0.16
	Manipur	1.05	0.69	1.60
	Mizoram	0.58	0.39	0.86
	Tripura	0.99	0.63	1.58
	Meghalaya	0.58	0.38	0.90
	Assam	0.22	0.13	0.35
	Gujarat	0.03	0.02	0.06

	Maharashtra	0.07	0.04	0.11
	Goa	0.06	0.03	0.11
	Andhra Pradesh	0.21	0.14	0.32
	Telangana	0.27	0.18	0.42
	Karnataka	0.16	0.10	0.24
	Kerala	0.26	0.17	0.39
	Tamil Nadu	0.27	0.19	0.38
	Puducherry	0.29	0.20	0.42

Odds ratios for bidi smoking obtained from multilevel multivariable logistic regression models

Covariates	Categories	Odds Ratio	95% CI	
Age		1.04	1.03	1.04
Sex	Male	0.04	0.04	0.05
	Female			
Wealth	Poorer			
	Poor	0.82	0.73	0.92
	Middle	0.67	0.57	0.78
	Rich	0.45	0.37	0.54
	Richer	0.28	0.23	0.35
Education	No formal education			
	Less than primary	0.79	0.69	0.91
	Primary less than secondary	0.48	0.43	0.55
	Secondary and above	0.18	0.15	0.21
Occupation	Unemployed			
	Labourer	2.45	2.02	2.96
	Housewife/ Retired/ Student	0.73	0.59	0.90
	Self	1.87	1.55	2.26
	Private	1.34	1.06	1.70
	Government	1.47	1.10	1.95
Area of residence	Urban			
	Rural	1.63	1.41	1.89
States	Jammu & Kashmir			
	Himachal Pradesh	3.24	2.19	4.78
	Punjab	1.18	0.77	1.82
	Chandigarh	2.49	1.59	3.90
	Uttarakhand	3.71	2.51	5.49
	Haryana	4.69	3.18	6.92
	Delhi	2.38	1.51	3.75
	Rajasthan	1.40	0.96	2.03
	Uttar Pradesh	1.32	0.90	1.94
	Chhattisgarh	0.53	0.32	0.86
	Madhya Pradesh	0.83	0.56	1.24
	West Bengal	1.98	1.36	2.89
	Jharkhand	0.27	0.15	0.50
	Odisha	0.37	0.22	0.63
	Bihar	0.32	0.21	0.50
	Sikkim	0.15	0.08	0.29
	Arunachal Pradesh	0.50	0.28	0.89
	Nagaland	1.26	0.77	2.07
	Manipur	0.26	0.13	0.53
	Mizoram	0.11	0.05	0.25
	Tripura	4.73	3.02	7.42
	Meghalaya	2.29	1.49	3.52
	Assam	0.83	0.54	1.26

	Gujarat	0.93	0.62	1.39
	Maharashtra	0.25	0.16	0.41
	Goa	0.27	0.14	0.50
	Andhra Pradesh	0.51	0.33	0.78
	Telangana	0.52	0.33	0.81
	Karnataka	0.61	0.40	0.91
	Kerala	0.42	0.25	0.69
	Tamil Nadu	0.62	0.41	0.93
	Puducherry	0.23	0.13	0.41

Odds ratios for smokeless tobacco use obtained from multilevel multivariable logistic regression models

Covariates	Categories	Odds Ratio	95% CI	
Age		1.02	1.02	1.02
Sex	Male			
	Female	0.44	0.41	0.46
Wealth	Poorer			
	Poor	0.88	0.83	0.94
	Middle	0.75	0.69	0.82
	Rich	0.56	0.50	0.62
	Richer	0.38	0.34	0.43
Education	No formal education			
	Less than primary	0.90	0.83	0.97
	Primary less than secondary	0.74	0.69	0.79
	Secondary and above	0.43	0.39	0.46
Occupation	Unemployed			
	Labourer	1.81	1.61	2.04
	Housewife/ Retired/ Student	0.67	0.60	0.76
	Self	1.46	1.30	1.64
	Private	1.56	1.36	1.79
	Government	0.99	0.84	1.17
Area of residence	Urban			
	Rural	1.09	1.01	1.19
States	Jammu & Kashmir			
	Himachal Pradesh	0.49	0.31	0.76
	Punjab	1.97	1.37	2.82
	Chandigarh	1.23	0.82	1.84
	Uttarakhand	2.35	1.66	3.34
	Haryana	1.39	0.95	2.03
	Delhi	2.64	1.83	3.82
	Rajasthan	3.03	2.19	4.19
	Uttar Pradesh	8.56	6.27	11.67
	Chhattisgarh	14.66	10.56	20.35
	Madhya Pradesh	7.32	5.34	10.04
	West Bengal	5.72	4.15	7.87
	Jharkhand	10.67	7.66	14.85
	Odisha	17.56	12.62	24.42
	Bihar	5.49	4.01	7.53
	Sikkim	2.21	1.51	3.23
	Arunachal Pradesh	11.81	8.29	16.82
	Nagaland	18.80	13.30	26.57
	Manipur	36.81	26.07	51.99
	Mizoram	17.94	12.69	25.36
	Tripura	40.00	28.21	56.70

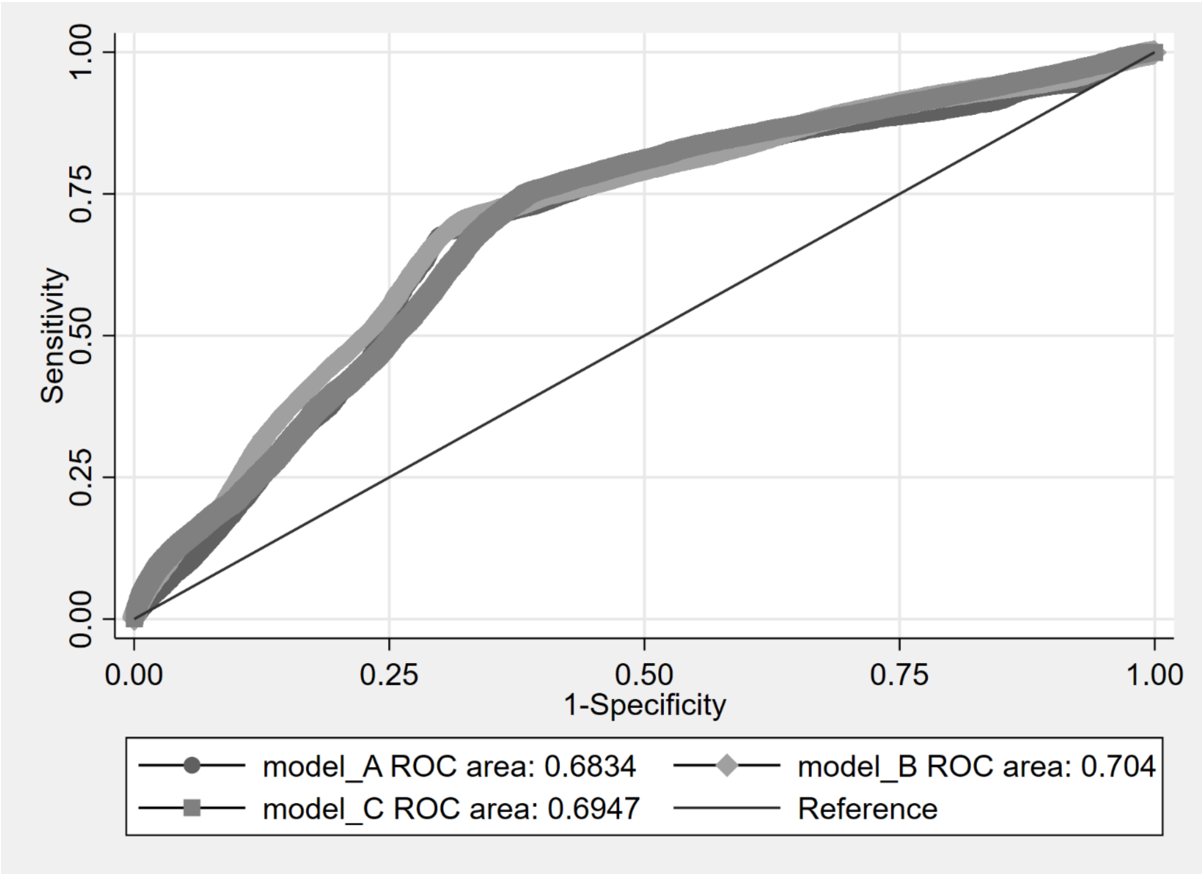
	Meghalaya	7.70	5.40	10.98
	Assam	17.68	12.92	24.20
	Gujarat	5.31	3.85	7.32
	Maharashtra	6.81	4.98	9.31
	Goa	1.76	1.21	2.56
	Andhra Pradesh	1.14	0.79	1.65
	Telangana	2.05	1.44	2.92
	Karnataka	3.63	2.62	5.02
	Kerala	1.12	0.76	1.65
	Tamil Nadu	2.15	1.54	3.00
	Puducherry	1.18	0.81	1.72

Odds ratios for dual use obtained from multilevel multivariable logistic regression models

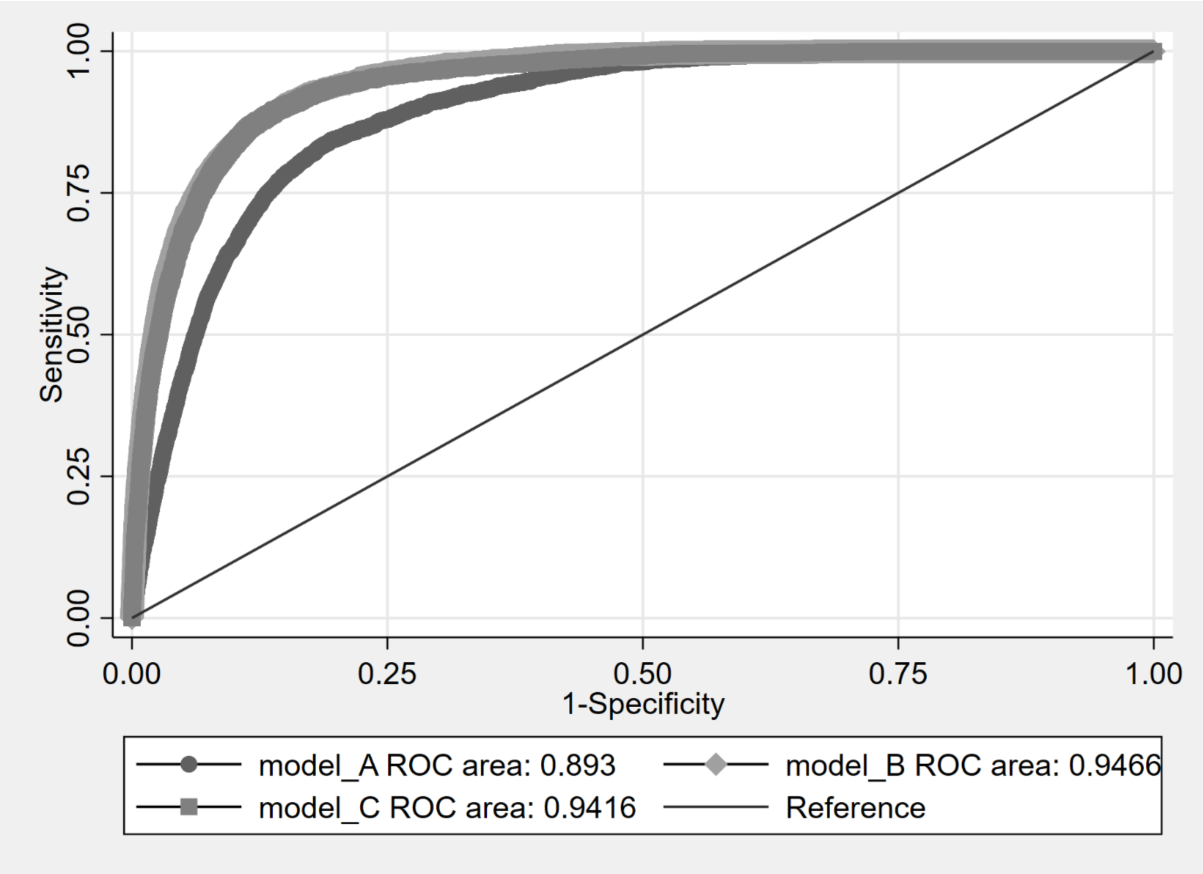
Covariates	Categories	Odds Ratio	95% CI	
Age		1.01	1.01	1.01
Sex	Male			
	Female	0.05	0.04	0.05
Wealth	Poorer			
	Poor	0.85	0.74	0.98
	Middle	0.74	0.61	0.89
	Rich	0.64	0.51	0.80
	Richer	0.45	0.35	0.58
Education	No formal education			
	Less than primary	0.82	0.69	0.98
	Primary less than secondary	0.62	0.54	0.73
	Secondary and above	0.26	0.22	0.31
Occupation	Unemployed			
	Labourer	2.18	1.72	2.77
	Housewife/ Retired/ Student	0.37	0.28	0.49
	Self	1.41	1.11	1.78
	Private	1.58	1.20	2.08
	Government	1.12	0.81	1.56
Area of residence	Urban			
	Rural	1.09	0.92	1.29
States	Jammu & Kashmir			
	Himachal Pradesh	1.00	0.49	2.04
	Punjab	1.37	0.70	2.68
	Chandigarh	1.34	0.65	2.74
	Uttarakhand	5.60	3.07	10.19
	Haryana	2.03	1.06	3.89
	Delhi	2.77	1.43	5.36
	Rajasthan	1.45	0.79	2.65
	Uttar Pradesh	7.29	4.17	12.74
	Chhattisgarh	2.06	1.06	3.99
	Madhya Pradesh	2.88	1.61	5.16
	West Bengal	3.06	1.70	5.52
	Jharkhand	7.41	4.04	13.57
	Odisha	4.55	2.44	8.48
	Bihar	1.73	0.95	3.14
	Sikkim	2.06	1.05	4.06
	Arunachal Pradesh	18.94	10.26	34.97
	Nagaland	17.87	9.68	33.00
	Manipur	28.55	15.51	52.53
	Mizoram	2.81	1.44	5.49
	Tripura	22.80	12.21	42.58
	Meghalaya	2.09	1.07	4.08
	Assam	7.16	4.04	12.69
	Gujarat	1.23	0.66	2.30

	Maharashtra	0.59	0.30	1.17
	Goa	0.30	0.11	0.81
	Andhra Pradesh	0.23	0.10	0.53
	Telangana	0.21	0.08	0.52
	Karnataka	1.33	0.72	2.46
	Kerala	1.19	0.59	2.37
	Tamil Nadu	0.65	0.34	1.27
	Puducherry	0.40	0.18	0.89

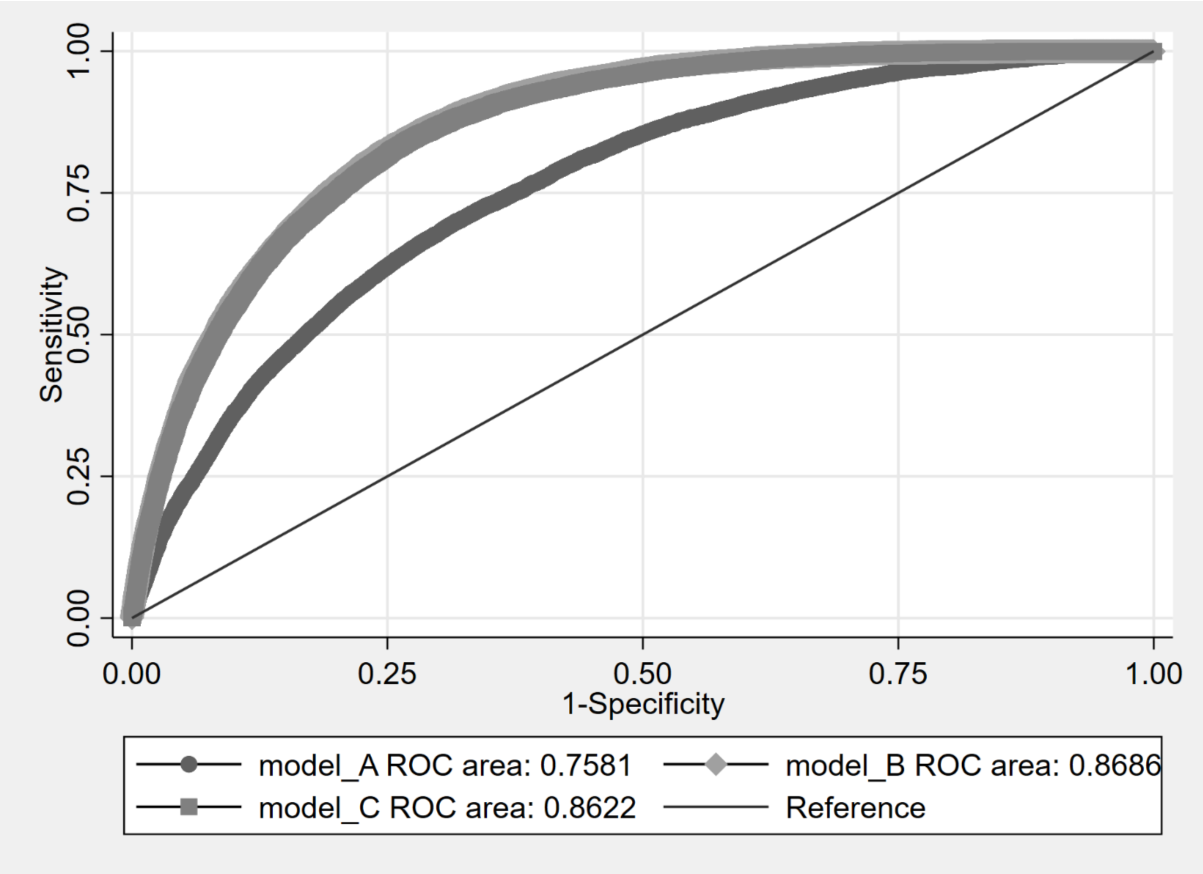
Area under the receiver operating characteristic (AU-ROC) curve for *cigarette smoking* plotted separately for single and multilevel logistic regression models



Area under the receiver operating characteristic (AU-ROC) curve for *bidi smoking* plotted separately for single and multilevel logistic regression models



Area under the receiver operating characteristic (AU-ROC) curve for *SLT use* plotted separately for single and multilevel logistic regression models



Area under the receiver operating characteristic (AU-ROC) curve for *dual use* plotted separately for single and multilevel logistic regression models

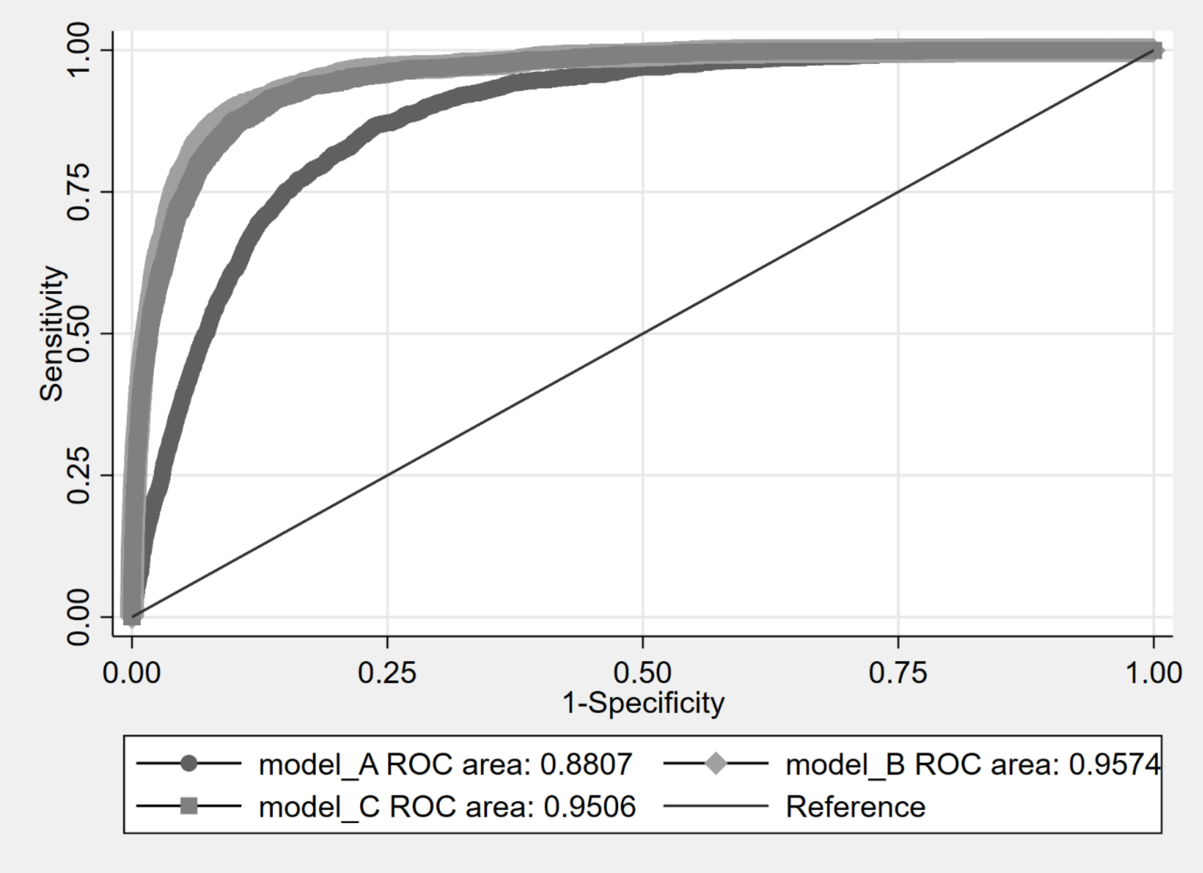


Table showing intraclass correlation coefficients obtained from three level hierarchical models with individual nested within city-wards/villages nested within states

		Null Model ICC 95% CI	Full model ICC 95% CI
Any tobacco use	State	17% (11%, 25%)	17% (11%, 25%)
	City ward/ village within state	23% (17%, 30%)	23% (17%, 30%)
Cigarette smoking	State	23% (15%, 34%)	22% (14%, 32%)
	City ward/ village within state	33% (25%, 42%)	33% (26%, 42%)
Bidi smoking	State	16% (11%, 25%)	19% (13%, 29%)
	City ward/ village within state	30% (24%, 36%)	33% (26%, 40%)
SLT use	State	28% (20%, 40%)	26% (18%, 36%)
	City ward/ village within state	36% (28%, 46%)	33% (25%, 42%)
Dual use	State	27% (18%, 38%)	26% (18%, 37%)
	City ward/ village within state	39% (31%, 48%)	43% (36%, 51%)

ICC: Intra-class correlation
Null model: Intercept only
Full model: Age, sex, education, occupation, wealth and area of residence

Interpretation

Example any tobacco use: Conditioned on covariates (age, sex, education, occupation, wealth and area of residence) any tobacco use is highly correlated within states (ICC: 17%). Within the same cityward/ village and state, this correlation was even higher (ICC: 23%).

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2-3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	5
Methods			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	6
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6
Bias	9	Describe any efforts to address potential sources of bias	6
Study size	10	Explain how the study size was arrived at	6
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	6-9
		(*b*) Describe any methods used to examine subgroups and interactions	6-9
		(*c*) Explain how missing data were addressed	6-9
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	6-9
		(*e*) Describe any sensitivity analyses	

Continued on next page

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	9
		(b) Give reasons for non-participation at each stage	
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	
		(b) Indicate number of participants with missing data for each variable of interest	
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	9-10
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	
		(*b*) Report category boundaries when continuous variables were categorized	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	
Discussion			
Key results	18	Summarise key results with reference to study objectives	10
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	10-11
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	11-12
Generalisability	21	Discuss the generalisability (external validity) of the study results	12
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	17

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Geographic variation in tobacco use in India: A population based multi-level cross-sectional study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-033178.R2
Article Type:	Original research
Date Submitted by the Author:	22-Apr-2020
Complete List of Authors:	Singh, Ankur; University of Melbourne, Centre for Health Equity, Melbourne School of Population & Global Health Arora, Monika; Public Health Foundation of India, Health Promotion and Tobacco Control Bentley, Rebecca; University of Melbourne, Centre for Health Equity, Melbourne School of Population & Global Health; University of Melbourne, Centre for Epidemiology and Biostatistics, Melbourne School of Population & Global Health Spittal, Matthew; The University of Melbourne, Melbourne School of Population and Global Health Do, Loc; The University of Adelaide, Australian Research Centre for Population Oral Health (ARCPOH), Adelaide Dental School Grills, Nathan; The University of Melbourne, Nossal Institute for Global Health English, Dallas; University of Melbourne, Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Smoking and tobacco
Keywords:	Tobacco, India, EPIDEMIOLOGY, PUBLIC HEALTH, SOCIAL MEDICINE

Geographic variation in tobacco use in India: A population based multi-level cross-sectional study

List of authors: Ankur Singh (Ph.D.),[1] Monika Arora (Ph.D.),[2] Rebecca Bentley (Ph.D.),[1,3] Matthew J Spittal (Ph.D.),[4] Loc G Do (Ph.D.),[5] Nathan Grills (Ph.D.),[6] Dallas R English (Ph.D)[3]

1. Centre for Health Equity, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Victoria, Australia
2. Health Promotion and Tobacco Control, Public Health Foundation of India, Gurugram, Haryana, India.
3. Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Victoria, Australia
4. Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Victoria, Australia
5. Australian Research Centre for Population Oral Health (ARCPOH), Adelaide Dental School, The University of Adelaide, Adelaide, South Australia, Australia
6. Nossal Institute for Global Health, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Victoria, Australia

Corresponding author:

Dr. Ankur Singh

Centre for Health Equity, Melbourne School of Population and Global Health,

The University of Melbourne,

Level 3, 207 Bouverie Street,

Melbourne, Victoria 3010, Australia

Tel: +61 3 8344 9256 Email: ankur.singh@unimelb.edu.au

Journal: BMJ Open

Publication category: Original Research

Running head: Area-level variations in tobacco use in India

Word count (excluding abstract, references, tables, and figures): 3158 words

Number of references: 43

Conflict of interest: None Declared

Abstract

Objective: This study aims to quantify the extent to which people's use of tobacco products varies by local areas (city-ward/village) across India and the variation in this clustering by tobacco product.

Design: Cross-sectional study

Setting and participants: Data on 73,954 adults across 2,547 city wards and villages was available for analysis from 30 states and two union territories in India.

Primary and secondary outcome measures: We included as primary outcomes self-reported any tobacco use, current cigarette smoking, current bidi smoking, current smokeless tobacco use and a derived variable for dual use describing respondents who engaged in both smoking and smokeless tobacco use.

Results: The median risk of an individual using tobacco was 1.64 times greater if a person hypothetically moved from an area of low to high risk of tobacco use (95% CI 1.60 – 1.69). Area-level partitioning of variation differed by tobacco product used. MORs ranged from 1.77 for smokeless tobacco use to 2.53 for dual use.

Conclusion: Tobacco use is highly clustered geographically in India. To be effective in India, policy interventions should be directed at the influence of specific local contextual factors on adult tobacco use. Where people live in India influences their use of tobacco, and this association may be greater than has been observed in other settings. Tailoring tobacco control policies for local areas in India may, therefore, provide substantial public health benefits.

Strengths and limitations of this study:

- First time application of multilevel analysis to quantify variations in tobacco use among local areas in a low- and middle- income country
- Multiple measures were estimated (ICC, MOR and discriminatory accuracy) that corroborated importance of local areas in determining tobacco use
- Large and nationally representative data on tobacco use was analysed
- Individual-level policy and economic variables were excluded to avoid atomistic fallacy
- Lacking area-level variables restricted analysis of their role in area variations in tobacco use

Introduction

Four fifths of the world's current smokers reside in low- and middle-income countries (LMICs), creating enormous societal and public health challenges. [1] The number of deaths from tobacco-related causes and loss of productivity is rapidly increasing in these, often resource poor, settings. [2 3]

The latest Indian Global Adult Tobacco Survey found that nearly 30% of all Indian adults use tobacco. [4] Additionally, the widespread use of smokeless tobacco presents a complex challenge for health systems and tobacco control because of its strong relationship with oral cancerous and pre-cancerous lesions. [5] Despite a nation-wide smokeless tobacco ban implemented in 2013-14, 20% of all tobacco users are smokeless tobacco users. [4] Added to this, the burden of tobacco use in India is disproportionally high among people who are socially disadvantaged. [6-8]

There is consistent evidence that local social and policy contexts shape patterns of tobacco use. [9] Multilevel studies (that simultaneously examine individual- and group-level determinants of health) from The Netherlands, Australia, South Africa, Mexico, Scotland, India, the USA and the UK suggest evidence of an association between area-level contextual factor (such as social disadvantage and local policy environments) and smoking.[10-22] For example, a study of Indian high school students from Mumbai reported the density of tobacco vendors around schools was associated with increased tobacco use by students.[23]

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]

To redress this important gap in evidence, this study aims to quantify the extent to which people's use of tobacco products varies by local areas (city-ward/village) across India and the variation in this clustering by tobacco product.

Methods

Study population

Data on tobacco use in India was obtained from the Global Adult Tobacco Survey (GATS 2) conducted in 2016 and 2017. GATS 2 is a multi-country household tobacco prevalence survey designed to support implementation of tobacco control within study countries.[4] Participants eligible for the survey were non-institutionalised individuals aged 15 years and older. The survey applied a multistage sampling procedure with different sampling hierarchies for urban and rural areas. For urban areas, city wards were the primary sampling unit from which census enumeration blocks, and then households, were selected. In rural areas, the primary sampling units were villages, from which households were selected. A total of 73,954 adults across 2,547 city wards and villages were available for analysis from 30 states and union territories in India. The response rate was 93%. [4]

Data collection

GATS-2 collected data using household and individual questionnaires developed in English and translated into 19 regional languages. The interviewer-administered questionnaires collected data on demographic characteristics, tobacco smoking, smokeless tobacco use, second hand smoke, socioeconomic position, media and knowledge, attitude and perceptions related to tobacco use. More details on sampling procedures and methods of data collection are published elsewhere. [4 30 31]

Outcomes

We included as primary outcomes self-reported any tobacco use, current cigarette smoking, current bidi smoking, current smokeless tobacco use and a derived variable for dual use describing respondents who engaged in both smoking and smokeless tobacco use. Participants were asked 'On average, how many of the following products do you currently smoke each day?. [4 30 31] We categorized those who reported smoking one or more manufactured/rolled tobacco in paper/leaf as current cigarette smokers. Similarly, we identified those who reported smoking one or more bidi as current bidi smokers. Regarding smokeless tobacco use, participants were asked 'Do you currently use smokeless tobacco on a daily basis, less than daily, or not at all? '. [4 30 31] We recorded those answering 'daily' or 'less than daily' as yes for current smokeless tobacco use. Those identified to be both current smokers (cigarette or bidi) and current smokeless tobacco users were identified as dual users. Therefore, we created five binary variables including any tobacco use, current cigarette smokers, current bidi smokers, current smokeless tobacco users and dual users.

Geographic level of aggregation (local areas)

Individuals from urban areas were clustered within city-wards and those in rural areas were clustered within villages. In urban areas, city wards are the units for local government

operations in India, responsible for essential community services including healthcare, education, housing, transport and so on.[32] In rural areas, villages make up the boundary for local panchayat (traditional local self-governance).[32]

Covariates

To account for compositional differences in populations within area-level clusters, we included individual-level demographic characteristics: age (as a continuous variable), sex and socioeconomic position (education: no formal education/less than primary/primary/ secondary or more; occupation: unemployed/labourer/housewife, retired, student/ self-employed/ private/ government; and household-level wealth: quintiles, 1 = lowest, 5 = highest) as covariates in the multilevel regression models. These variables were selected based on a previous study.[7]

Statistical Analysis

We performed the statistical analyses using Stata 15.0 (Statacorp, College Station, TX, USA). We used survey commands to account for the complex survey design and to perform the weighted descriptive analysis. We plotted the prevalence and 95% confidence intervals for any tobacco use and for different types of tobacco use to visually examine their variation by local areas. We fitted multilevel logistic regression models with random intercepts for local areas and fixed slopes with individuals nested in city wards or villages respectively. Multilevel models operationalise studying population-level variations in health outcomes by examining the extent of clustering in health outcomes that exists at the group or contextual level.[24 33-37] Using intra-class correlation coefficients (ICC) and median odds ratios (MOR), we decomposed the variance in tobacco use at city-ward or village levels. The ICC is expressed as a percentage and is interpreted in these analyses as the share of inter-individual variation in health outcome that exists at the group level. For example, an ICC of 8% at the village level means that of all the individual-level variation in tobacco use among rural areas, 8% is attributed to the village level. The higher the individual correlation in health outcomes within a context, the more relevant is the context for understanding individual differences in the health outcome. [25] We estimated the MOR, which describes the area-level variance as an odds ratio, as the median value of the distribution of odds ratios obtained when two individuals with the same covariate values are picked from two different areas, comparing the one from the higher prevalence area to the one from the area with lower prevalence. [24 36 38] In the absence of any area-level variation, the MOR is equal to one. We estimated both MORs and ICCs for binary outcomes as the partition of variance between different levels does not have the same intuitive interpretation as a linear model. [24 38] We estimated ICCs and MORs from intercept only models to examine presence of clustering, and heterogeneity between areas, in the outcomes of tobacco use.

We also applied an alternate method to examine the relevance of area-level contexts for tobacco use by comparing discriminatory accuracies obtained from fitted single-level and multi-level logistic regression models. The area under the receiver operating characteristic curve (AU-ROC) was constructed by plotting the true positive fraction (TPF, sensitivity) against the false positive fraction (FPF, 1 – specificity). It measures the ability of the model to classify individuals with and without the outcome and takes a value between 0.5 and 1.0 where 1.0 is perfect discrimination and 0.5 where covariates have no predictive power. [24]

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 assessed goodness of fit by estimating the changes in the Deviance Information Criterion (DIC). All models were fitted separately for each type of tobacco use (cigarette smoking, bidi smoking, smokeless tobacco use and dual use) to determine any differences in variations in tobacco use according to different types of tobacco use. We performed a sensitivity analysis to examine clustering in tobacco use in city/wards and villages within states by fitting three-level hierarchical models: individual nested within city/wards and villages nested within states.

Patient and Public Involvement

No patients or public were involved in this study.

Results

We analysed data for 73,954 individuals (99.9%) of the 74,037 survey participants. We did not analyse data on 83 participants due to missing covariates data. Table 1 shows descriptive characteristics of the sample according to residence status. 28% of adults used tobacco products. The prevalence of smokeless tobacco use was 18.6% (Table 1). Plots for prevalence and 95% CI for any tobacco use and different types of tobacco use by local areas showed substantial variations (Supplementary appendix).

Intercept only models (null models with no covariate adjustment) estimated 22% (95% CI: 20, 24) of any tobacco use was clustered at the city-ward/village level. Cigarette smoking was clustered 31%, bidi smoking at 28%, dual use at 40% and smokeless tobacco at 36% respectively (estimates not reported in the tables). For each outcome, the AUC increased when multilevel logistic regression models were fitted. The AUC increased to 0.86 in Model B as compared to 0.79 in a 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).

After including all individual-level covariates, the proportion of variance attributable to the areas remained at 21% (95% CI: 20, 22) (Table 2). Correspondingly, the median odds ratio for was 2.43 (95% CI: 2.35, 2.52). 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. Including area of residence and state in Model C substantially reduced the estimates of proportion of variance attributable to areas and the respective median odds ratios. The proportion of variance for any tobacco use reduced from 21% to 7.6% and corresponding median ratio from 2.42 to 1.64. Sensitivity analysis confirmed our findings of high clustering in any tobacco use within city-ward or villages from the same state (Supplementary Appendix pp.21).

The decrease in DIC values between the single-level models and multilevel models including covariates suggested better model fit (Table 2).

Among the different types of tobacco use, the highest ICC (22%; 95% CI: 19, 26) and MOR (2.53; 95% CI: 2.32, 2.74) were for dual use and the lowest for SLT use ((ICC: 10%; 95% CI: 9, 11), (MOR: 1.77; 95% CI: 1.71, 1.83)) (Table 3). Similar to any tobacco use, substantial reductions in estimates of ICC and MOR were observed upon inclusion of state and area of residence in Model C compared with Model B.

Discussion

We found substantial variation in tobacco use across local areas in India. Individual-level social and demographic characteristics were not able to explain the high area-level variations in tobacco. Including states and area of residence explained substantial area-level variation in tobacco use. However, the remaining variation in tobacco use was still high, indicating the importance of local areas. The degree of area-level variation in tobacco use differed according to the types of tobacco product. Dual use (smoking and smokeless) had the highest geographic clustering.

Strengths and Limitations

This study had several strengths and limitations. To the best of our knowledge this is the first study from LMICs that has studied variation in tobacco use at local area level using a nationwide representative data. [4] By using different measures (ICC, MOR and AUC) we not only inform the extent of variation but we comprehensively examine the degree of clustering, the heterogeneity in outcomes among areas as well as the ability of local areas to classify individuals according to tobacco use. [24 33-36] This study also has limitations. 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 the respondent's perceptions rather than objective measures of availability and implementation of policies in local areas and because this data was only gathered from smokers. The non-ecologic nature of these variables could lead to falsely attributing individual-level measures to area levels (the atomistic fallacy). [37]

Discussion in context of current evidence

Our findings of high variations in tobacco use among local areas is new. A multilevel study on societal determinants of tobacco use from Scotland found no evidence of clustering in tobacco use at the area level. [18] Other multilevel studies have not presented measures of variance, which limits comparisons.[10-22] 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.

Tobacco specific variations in the values of ICC and MOR highlight potential differences in the relevance of contexts by type of tobacco product used. Evidence from other studies suggests that while wealthier and more educated individuals have higher odds of cigarette smoking than their disadvantaged counterparts, disadvantaged individuals have higher odds of bidi smoking and smokeless tobacco use. [6 7] Our study highlights the presence of both individual and geographic socioeconomic inequalities in tobacco use by product. For example, we observed a higher effect of individual social and demographic characteristics in smokeless tobacco use when compared to cigarette smoking and bidi smoking for contextual effects (change in ICC from 36% in null model to 31% in adjusted model). In addition, the proportion of variation for all types of tobacco use was markedly explained by adding states into the model. This emphasizes the role of cultural and regional diversity within India in determining tobacco use.[39] Both ICCs from the three-level hierarchical models and odds ratios estimated from regression models confirmed pivotal role played by states in geographic inequities in tobacco use in India (see Supplementary Appendix).

Research and policy implications

Given the role of contexts in shaping individual health behaviours, this study builds a framework for operationalizing a contextual thinking in tobacco control activities, particularly in LMICs where social norms and cultural aspects may differ from high-income countries. High general contextual effects of local areas for tobacco use necessitates a thorough examination of factors at the area-level that may be causally associated with individual tobacco use as well as those which can explain the high variations in tobacco use among local areas. This may only be possible if either data on individual-level tobacco use is linked with small area characteristics, or if future population-based surveys collect both area- and individual-level data relevant to tobacco use. 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. Furthermore, current findings build the platform for more robust population-based studies that collectively examine area- and individual-level determinants of tobacco use in India and other LMICs.

This study has several policy implications. Our findings confirm that context plays an important role in determining use of tobacco. India's Cigarettes and other Tobacco Products Act (COTPA) is a national law, which is in line with World Health Organization's Framework Convention on Tobacco Control. States at sub-national level are responsible for implementing various tobacco control policy measures under COTPA. Comparison of GATS-2 and GATS-1, and household surveys, has highlighted changes in prevalence of tobacco use due to differential implementation of these measures.[40 41] States are also allowed to develop context specific information, education and communication resources to match the local needs. [41 42] Therefore, health promotion and tobacco control interventions must be designed for contexts and applied contextually rather than being individually oriented.[9 43] There is the potential to enhance National Tobacco Control Program's (NTCP) implementation at city-ward, village and block level as well.[41] 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. 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 disadvantange to reduce geographic inequalities.

Conclusion

Where people live in India influences their use of tobacco, and this association may be greater than has been observed in other settings. Tailoring tobacco control policies for local areas in India may, therefore, provide substantial public health benefits.

References

1. WHO. Tobacco: Fact sheet Geneva: World Health Organization; 2018 [Available from: https://www.who.int/en/news-room/fact-sheets/detail/tobacco accessed 10 December 2018.
2. Hipple Walters B, Petrea I, Lando H. Tobacco Control in Low- and Middle-Income Countries: Changing the Present to Help the Future. *Journal of Smoking Cessation* 2018:1-2. doi: 10.1017/jsc.2018.4 [published Online First: 03/14]
3. Gilmore AB, Fooks G, Drope J, et al. Exposing and addressing tobacco industry conduct in low-income and middle-income countries. *The Lancet* 2015;385(9972):1029-43. doi: 10.1016/S0140-6736(15)60312-9
4. Tata Institute of Social Sciences (TISS) MaMoHaFW, Government of India. Global Adult Tobacco Survey GATS 2 India 2016-17. New Delhi: Tata Institute of Social Sciences (TISS), Mumbai and Ministry of Health and Family Welfare, Government of India, 2018.
5. Gupta PC, Arora M, Sinha D, et al. Smokeless tobacco and public health in India. *Ministry of Health & Family Welfare, Government of India; New Delhi* 2016
6. Corsi DJ, Subramanian SV. Divergent socio-economic gradients in smoking by type of tobacco use in India. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease* 2014;18(1):122-4. doi: 10.5588/ijtld.13.0246 [published Online First: 2013/12/25]
7. Singh A, Arora M, English DR, et al. Socioeconomic Gradients in Different Types of Tobacco Use in India: Evidence from Global Adult Tobacco Survey 2009-10. *Biomed Res Int* 2015;2015:837804. doi: 10.1155/2015/837804 [published Online First: 2015/08/15]
8. Bhan N, Srivastava S, Agrawal S, et al. Are socioeconomic disparities in tobacco consumption increasing in India? A repeated cross-sectional multilevel analysis. *BMJ open* 2012;2(5) doi: 10.1136/bmjopen-2012-001348 [published Online First: 2012/10/02]
9. Link BG, Phelan J. Social Conditions as Fundamental Causes of Disease. *Journal of Health and Social Behavior* 1995;35:80-94. doi: Doi 10.2307/2626958
10. Rachele JN, Wood L, Nathan A, et al. Neighbourhood disadvantage and smoking: Examining the role of neighbourhood-level psychosocial characteristics. *Health Place* 2016;40:98-105. doi: 10.1016/j.healthplace.2016.04.012 [published Online First: 2016/05/27]
11. Ellaway A, Macintyre S. Are perceived neighbourhood problems associated with the likelihood of smoking? *Journal of epidemiology and community health* 2009;63(1):78-80. doi: 10.1136/jech.2007.068767 [published Online First: 2008/12/18]
12. Morris T, Manley D, Van Ham M. Context or composition: How does neighbourhood deprivation impact upon adolescent smoking behaviour? *PloS one* 2018;13(2):e0192566. doi: 10.1371/journal.pone.0192566 [published Online First: 2018/02/09]
13. Mayne SL, Auchincloss AH, Moore KA, et al. Cross-sectional and longitudinal associations of neighbourhood social environment and smoking behaviour: the multiethnic study of atherosclerosis. *Journal of epidemiology and community health* 2017;71(4):396-403. doi: 10.1136/jech-2016-207990 [published Online First: 2016/11/26]
14. Chuang YC, Cubbin C, Ahn D, et al. Effects of neighbourhood socioeconomic status and convenience store concentration on individual level smoking. *Journal of*

epidemiology and community health 2005;59(7):568-73. doi: 10.1136/jech.2004.029041 [published Online First: 2005/06/21]
15. Fleischer NL, Lozano P, Arillo Santillan E, et al. The impact of neighbourhood violence and social cohesion on smoking behaviours among a cohort of smokers in Mexico. *Journal of epidemiology and community health* 2015;69(11):1083-90. doi: 10.1136/jech-2014-205115 [published Online First: 2015/06/06]
16. Turrell G, Hewitt BA, Miller SA. The influence of neighbourhood disadvantage on smoking cessation and its contribution to inequalities in smoking status. *Drug Alcohol Rev* 2012;31(5):645-52. doi: 10.1111/j.1465-3362.2012.00452.x [published Online First: 2012/04/18]
17. van Lenthe FJ, Mackenbach JP. Neighbourhood and individual socioeconomic inequalities in smoking: the role of physical neighbourhood stressors. *Journal of epidemiology and community health* 2006;60(8):699-705. doi: 10.1136/jech.2005.043851 [published Online First: 2006/07/15]
18. Shareck M, Ellaway A. Neighbourhood crime and smoking: the role of objective and perceived crime measures. *BMC public health* 2011;11:930. doi: 10.1186/1471-2458-11-930 [published Online First: 2011/12/16]
19. Fleischer NL, Thrasher JF, Saenz de Miera Juarez B, et al. Neighbourhood deprivation and smoking and quit behaviour among smokers in Mexico: findings from the ITC Mexico Survey. *Tobacco control* 2015;24 Suppl 3:iii56-iii63. doi: 10.1136/tobaccocontrol-2013-051495 [published Online First: 2014/08/30]
20. Lau YK, Tam J, Fleischer NL, et al. Neighbourhood deprivation, smoking, and race in South Africa: A cross-sectional analysis. *Prev Med Rep* 2018;11:202-08. doi: 10.1016/j.pmedr.2018.07.001 [published Online First: 2018/07/13]
21. Timmermans EJ, Veldhuizen EM, Snijder MB, et al. Neighbourhood safety and smoking in population subgroups: The HELIUS study. *Preventive medicine* 2018;112:111-18. doi: 10.1016/j.ypmed.2018.04.012 [published Online First: 2018/04/15]
22. Reijneveld SA. Neighbourhood socioeconomic context and self reported health and smoking: a secondary analysis of data on seven cities. *Journal of epidemiology and community health* 2002;56(12):935-42. [published Online First: 2002/12/04]
23. Mistry R, Pednekar M, Pimple S, et al. Banning tobacco sales and advertisements near educational institutions may reduce students' tobacco use risk: evidence from Mumbai, India. *Tobacco control* 2015;24(e1):e100-7. doi: 10.1136/tobaccocontrol-2012-050819 [published Online First: 2013/08/21]
24. Merlo J, Wagner P, Ghith N, et al. An Original Stepwise Multilevel Logistic Regression Analysis of Discriminatory Accuracy: The Case of Neighbourhoods and Health. *PloS one* 2016;11(4):e0153778. doi: 10.1371/journal.pone.0153778 [published Online First: 2016/04/28]
25. Merlo J, Wagner P, Austin PC, et al. General and specific contextual effects in multilevel regression analyses and their paradoxical relationship: A conceptual tutorial. *SSM - Population Health* 2018;5:33-37. doi: 10.1016/j.ssmph.2018.05.006
26. Diez Roux AV. The study of group-level factors in epidemiology: rethinking variables, study designs, and analytical approaches. *Epidemiologic reviews* 2004;26(1):104-11. doi: 10.1093/epirev/mxh006 [published Online First: 2004/07/06]
27. Merlo J. Multilevel analytical approaches in social epidemiology: measures of health variation compared with traditional measures of association. *Journal of epidemiology and community health* 2003;57(8):550-2. [published Online First: 2003/07/29]

28. Merlo J, Chaix B, Yang M, et al. A brief conceptual tutorial on multilevel analysis in social epidemiology: interpreting neighbourhood differences and the effect of neighbourhood characteristics on individual health. *Journal of epidemiology and community health* 2005;59(12):1022-8. doi: 10.1136/jech.2004.028035 [published Online First: 2005/11/16]
29. Merlo J, Viciana-Fernandez FJ, Ramiro-Farinas D, et al. Bringing the individual back to small-area variation studies: a multilevel analysis of all-cause mortality in Andalusia, Spain. *Soc Sci Med* 2012;75(8):1477-87. doi: 10.1016/j.socscimed.2012.06.004 [published Online First: 2012/07/17]
30. Palipudi KM, Morton J, Hsia J, et al. Methodology of the Global Adult Tobacco Survey - 2008-2010. *Global health promotion* 2016;23(2 Suppl):3-23. doi: 10.1177/1757975913499800 [published Online First: 2013/09/18]
31. GATS. Global Adult Tobacco Survey (GATS) India Report 2009-2010. : Ministry of Health and Family Welfare, India, 2010.
32. Office of the Registrar General & Census Commissioner I. Administrative Divisions New Delhi, India: Ministry of Home Affairs, Government of India; 2018 [Available from: http://censusindia.gov.in/Census_And_You/Administrative_division.aspx accessed 30/11/18 2018.
33. Subramanian SV. The relevance of multilevel statistical methods for identifying causal neighborhood effects. *Soc Sci Med* 2004;58(10):1961-7. doi: 10.1016/S0277-9536(03)00415-5 [published Online First: 2004/03/17]
34. Subramanian SV, Jones K, Kaddour A, et al. Revisiting Robinson: the perils of individualistic and ecologic fallacy. *International journal of epidemiology* 2009;38(2):342-60; author reply 70-3. doi: 10.1093/ije/dyn359 [published Online First: 2009/01/31]
35. Merlo J, Chaix B, Yang M, et al. A brief conceptual tutorial of multilevel analysis in social epidemiology: linking the statistical concept of clustering to the idea of contextual phenomenon. *Journal of epidemiology and community health* 2005;59(6):443-9. doi: 10.1136/jech.2004.023473 [published Online First: 2005/05/25]
36. Merlo J, Ohlsson H, Lynch KF, et al. Individual and collective bodies: using measures of variance and association in contextual epidemiology. *Journal of epidemiology and community health* 2009;63(12):1043-8. doi: 10.1136/jech.2009.088310 [published Online First: 2009/08/12]
37. Singh A, Harford J, Peres MA. Investigating societal determinants of oral health-Opportunities and challenges in multilevel studies. *Community dentistry and oral epidemiology* 2018;46(4):317-27. doi: 10.1111/cdoe.12369 [published Online First: 2018/02/21]
38. Merlo J, Chaix B, Ohlsson H, et al. A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *Journal of epidemiology and community health* 2006;60(4):290-7. doi: 10.1136/jech.2004.029454 [published Online First: 2006/03/16]
39. Sarkar BK, Reddy KS. Priorities for tobacco control research in India. *Addiction* 2012;107(12):2066-68. doi: 10.1111/j.1360-0443.2012.03942.x
40. 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):928-33. doi: 10.15585/mmwr.mm6841a1 [published Online First: 2019/10/18]

41. 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):103-10. doi: 10.1136/tobaccocontrol-2018-054621 [published Online First: 2018/12/17]

42. 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.

43. Phelan JC, Link BG, Tehranifar P. Social conditions as fundamental causes of health inequalities: theory, evidence, and policy implications. *J Health Soc Behav* 2010;51 Suppl(1_suppl):S28-40. doi: 10.1177/0022146510383498 [published Online First: 2010/12/22]

Footnotes

Author's contribution:
AS conceptualised the study, acquisition of data, analysed the data, interpretation of results, led the manuscript preparation and the submission process
MA contributed to interpretation of local policy implications of the results and drafts of the manuscript
RB and NG contributed by critical inputs on multiple draft of the manuscript and interpretation of results
MS and LD contributed by development of analytical framework, interpretation of results and revision of manuscript
DE contributed by reviewing early and advanced drafts of the manuscript, development of analysis plan and interpretation of results.

Funding statement: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. Matthew Spittal is a recipient of an Australian Research Council Future Fellowship (project number FT180100075) funded by the Australian Government
Competing interests: None declared.
Declaration of Interest: Authors have no conflict of interest to declare.
Patient consent for publication: Not required.
Data sharing statement: De-identified data is available from GTSS: https://www.cdc.gov/tobacco/global/gtss/index.htm
Ethics approval: No formal ethical approval was required.
Provenance and peer review: Not commissioned; externally peer reviewed.

Tables and figures

Table 1. Descriptive characteristics of the sample (n=73,954)

Variable	Categories	Percentage
Age (years)	15 to 30	41.7
	31 to 45	29.7
	46 to 60	17.7
	61 to 75	8.9
	76 and above	1.8
Sex	Male	51.1
	Female	48.9
Wealth	Poorer	23.4
	Poor	36.5
	Middle	15.0
	Rich	12.2
	Richer	12.9
Education	No formal education	26.4
	Less than primary	9.2
	Primary	28.2
	Secondary or more	36.2
Occupation	Unemployed	4.3
	Labour	21.2
	Housewife/ Retired/ Student	44.1
	Self	19.4
	Private	8.3
	Government job	2.7
Area of residence	Urban	34.5
	Rural	65.5
Tobacco use	Non- user	72.2
	Cigarette smoking	1.3
	Bidi smoking	4.6
	Smokeless tobacco use	18.6
	Dual Use (Smokeless tobacco use + Smoking)	2.8
	Bidi + Cigarette	0.5

Weighted percentages (Using Survey Weights)

Table 2. Multilevel logistic regression models for any tobacco use among Indian adults (n=73,954 individuals nested in 2547 city wards and villages).

	Model A Estimate 95% CI	Model B Estimate 95% CI	Model C Estimate 95% CI
AUC	0.79	0.86	0.86
AUC change		0.07	0
Variance		0.87	0.27
		(0.80, 0.94)	(0.24, 0.30)
ICC		21%	8%
		(20, 22)	(7, 9)
MOR		2.43	1.64
		(2.35, 2.52)	(1.60, 1.69)
DIC	71171.7	66619.6	64702.3
DIC change		-4552.1	-1917.3

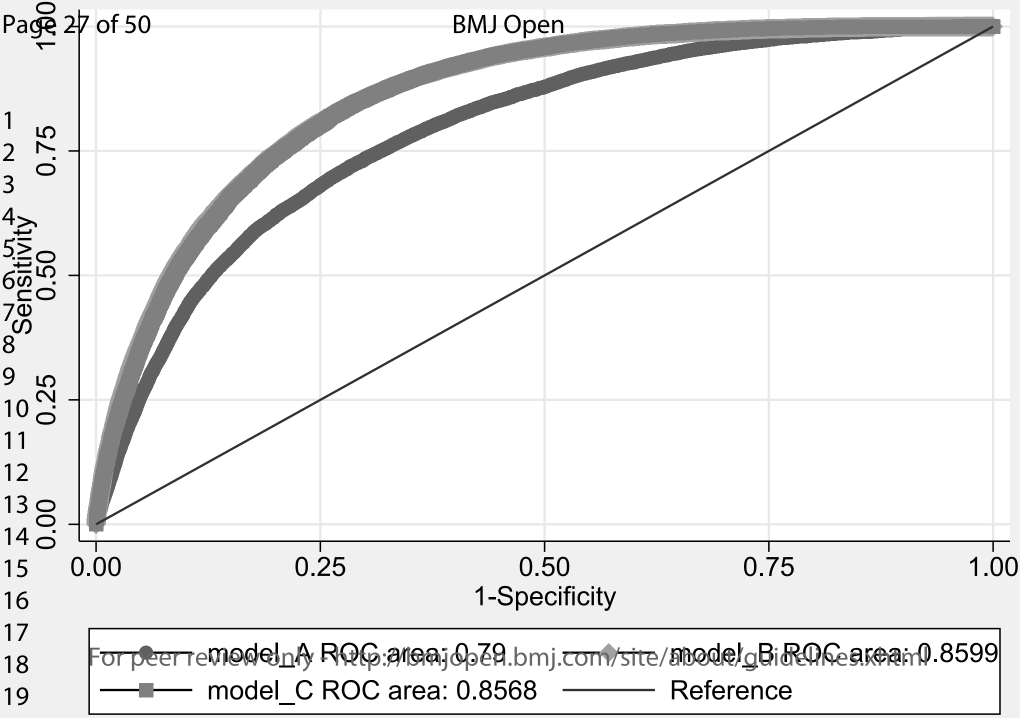
Model A: Single-level logistic regression model (Covariates included: age, sex, area of residence, education, occupation, wealth); Model B: Multi-level logistic regression model (Covariates included: age, sex, education, occupation, wealth); Model C: Multi-level logistic regression model (Covariates included: age, sex, education, occupation, wealth, area of residence and states). Reference group: No tobacco use.

Table 3. Multilevel logistic regression models for different types of tobacco use among Indian adults (n=73,954 individuals nested in 2547 city wards and villages).

	Model A Estimate 95% CI	Model B Estimate 95% CI	Model C Estimate 95% CI
Cigarettes (n=54,648)			
Variance		1.44 (1.19, 1.70)	0.53 (0.38, 0.68)
ICC (%)		30 (27, 34)	14 (11, 18)
MOR		3.14 (2.83, 3.46)	2.00 (2.82, 3.46)
DIC	10630.4	10175.0	9480.5
DIC change		-455.4	-694.5
AUC	0.68	0.70	0.69
AUC change		0.02	-0.01
Bidi (n=56,814)			
Variance		1.53 (1.33, 1.72)	0.65 (0.53, 0.76)
ICC (%)		32 (29, 35)	16 (14,19)
MOR		3.25 (3.01, 3.49)	2.15 (2.01, 2.30)
DIC	18822.5	17680.8	16765.4
DIC change		-1141.7	-915.4
AUC	0.89	0.95	0.94
AUC change		0.06	-0.01
SLT (n=66,089)			
Variance		1.46 (1.34, 1.59)	0.36 (0.31, 0.40)
ICC (%)		31 (29, 33)	10 (9, 11)
MOR		3.17 (3.01, 3.32)	1.77 (1.71, 1.83)
DIC	56207.3	51179.1	48915.1
DIC change		-5028.1	-2264.0
AUC	0.76	0.87	0.86
AUC change		0.11	-0.01
Dual use (n=55,522)			
Variance		2.41 (2.09, 2.72)	0.95 (0.78, 1.12)
ICC (%)		42 (39, 45)	22 (19, 26)
MOR		4.39 (3.96, 4.82)	2.53 (2.32, 2.74)
DIC	14335.7	12989.8	12045.9
DIC change		-1345.9	-943.9
AUC	0.88	0.96	0.95
AUC change		0.08	-0.01

Model A: Single-level logistic regression model (Covariates included: age, sex, area of residence, education, occupation, wealth); Model B: Multi-level logistic regression model (Covariates included: age, sex, education, occupation, wealth); Model C: Multi-level logistic regression model (Covariates included: age, sex, education, occupation, wealth, area of residence and states). Reference group: No tobacco use.

Figure 1. Area under the receiver operating characteristic (AU-ROC) curve for tobacco use plotted separately for single and multilevel logistic regression models


Sensitivity
0.00
0.25
0.50
0.75
1.00
0.00
0.25
0.50
0.75
1.00
1-Specificity
model_A ROC area: 0.79
model_B ROC area: 0.8599
model_C ROC area: 0.8568
Reference

Supplementary Appendix

Contents

Prevalence and 95% confidence intervals of any tobacco use by local areas 2

Prevalence and 95% confidence intervals of cigarette smoking by local areas 3

Prevalence and 95% confidence intervals of bidi smoking by local areas 4

Prevalence and 95% confidence intervals of SLT use by local areas 5

Prevalence and 95% confidence intervals of Dual use by local areas 6

Odds ratios for any tobacco use obtained from multilevel multivariable logistic regression models ... 7

Odds ratios for cigarette smoking obtained from multilevel multivariable logistic regression models 9

Odds ratios for bidi smoking obtained from multilevel multivariable logistic regression models 11

Odds ratios for smokeless tobacco use obtained from multilevel multivariable logistic regression models 13

Odds ratios for dual use obtained from multilevel multivariable logistic regression models 15

Area under the receiver operating characteristic (AU-ROC) curve for *cigarette smoking* plotted separately for single and multilevel logistic regression models 17

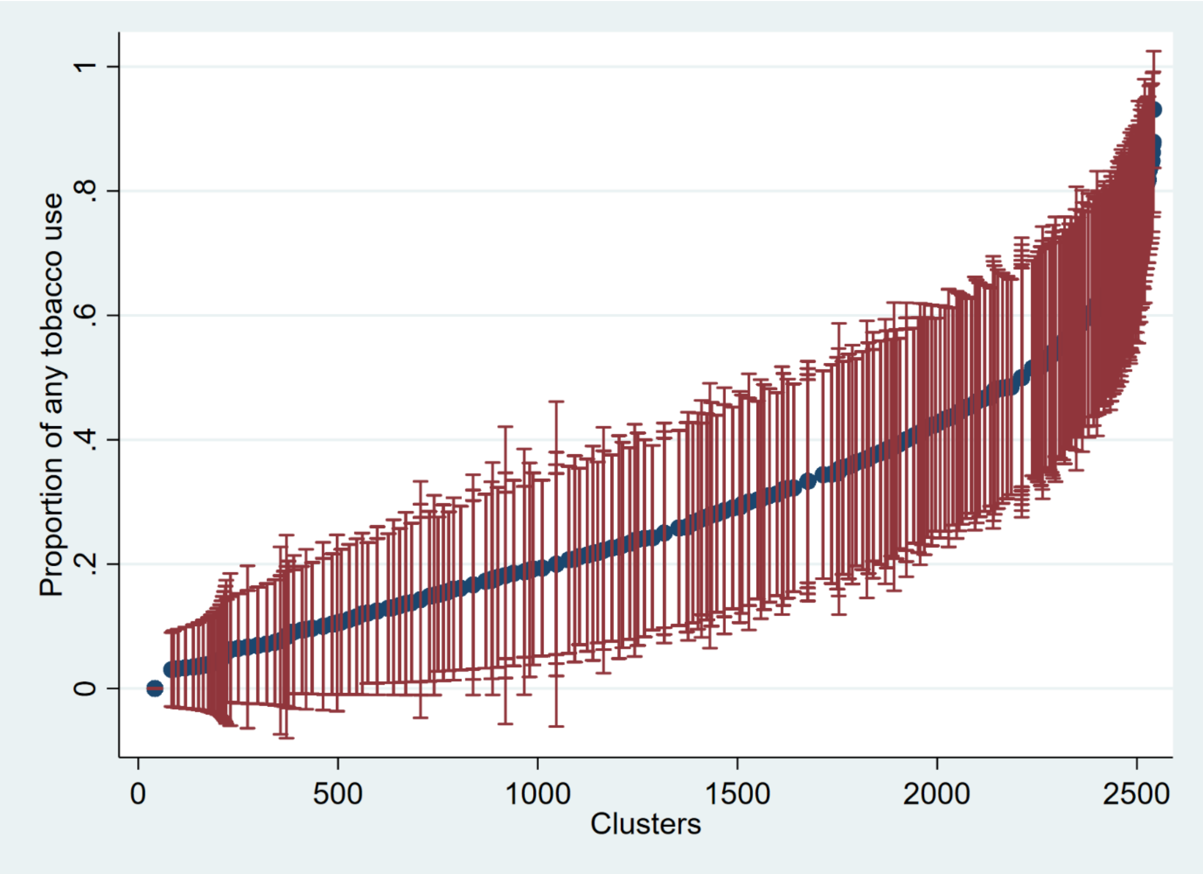
Area under the receiver operating characteristic (AU-ROC) curve for *bidi smoking* plotted separately for single and multilevel logistic regression models 18

Area under the receiver operating characteristic (AU-ROC) curve for *SLT use* plotted separately for single and multilevel logistic regression models 19

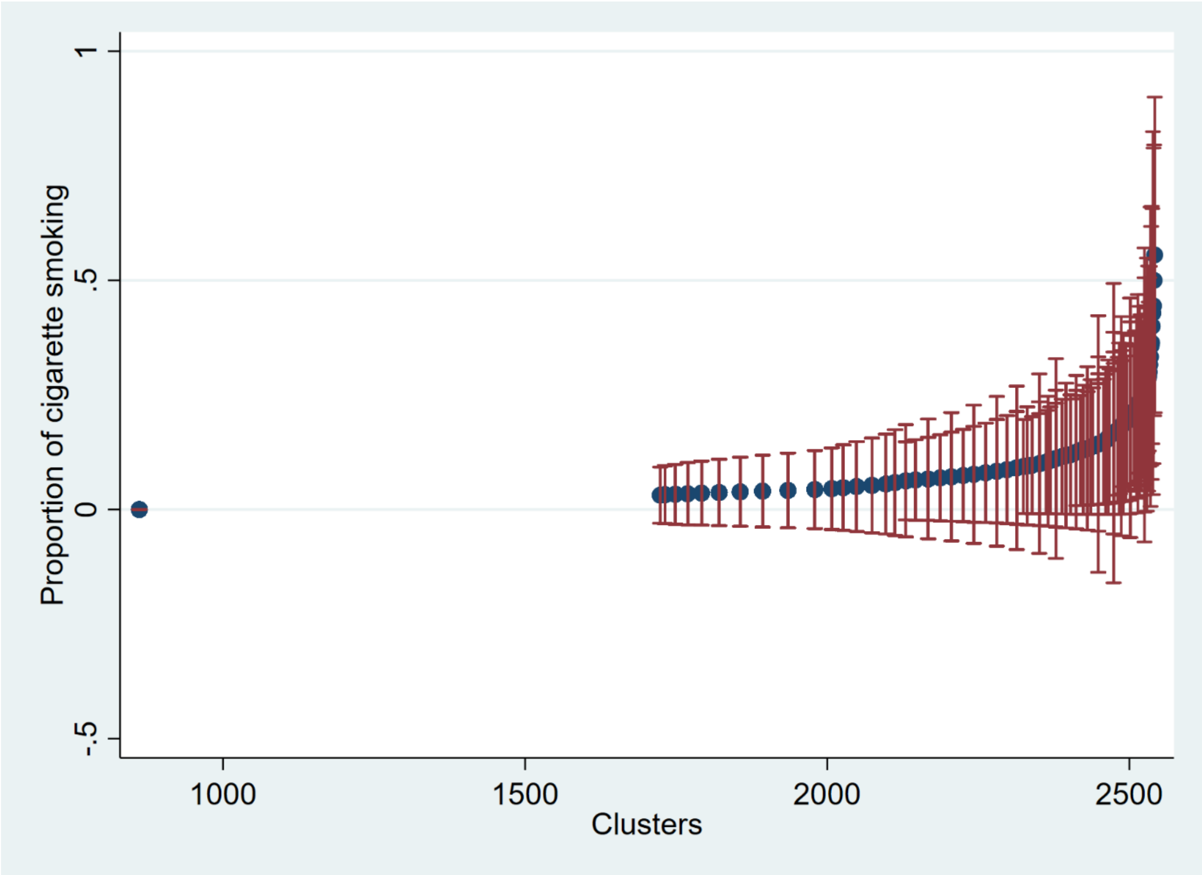
Area under the receiver operating characteristic (AU-ROC) curve for *dual use* plotted separately for single and multilevel logistic regression models 20

Table showing intraclass correlation coefficients obtained from three level hierarchical models with individual nested within city-wards/villages nested within states 21

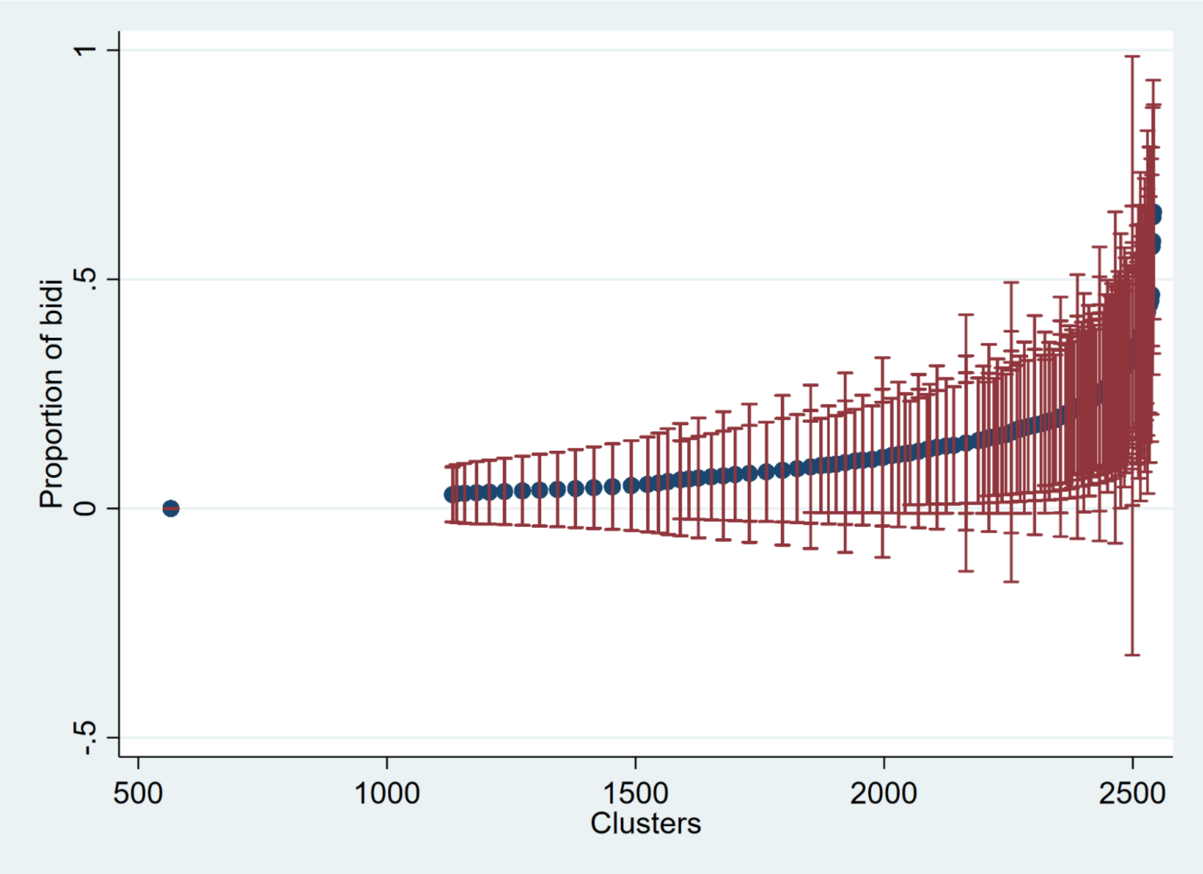
Prevalence and 95% confidence intervals of any tobacco use by local areas



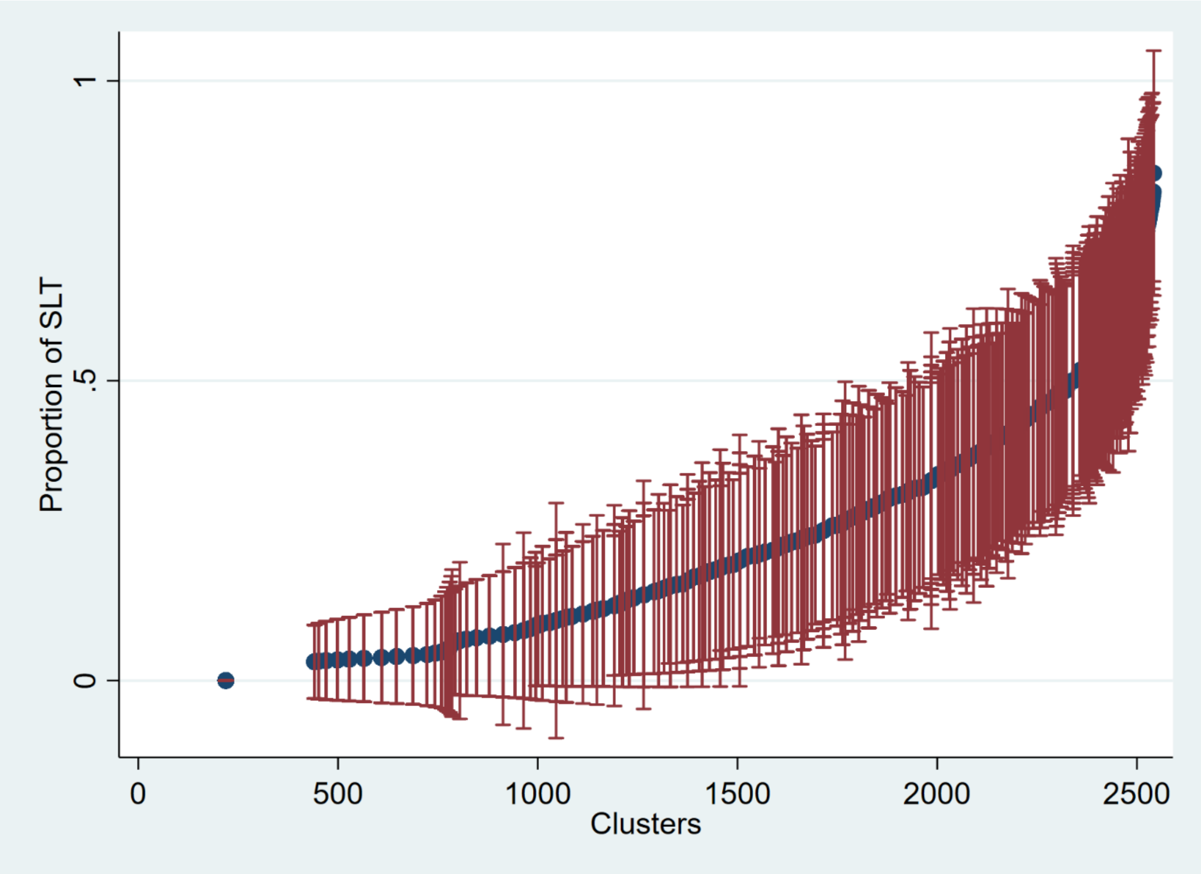
Prevalence and 95% confidence intervals of cigarette smoking by local areas



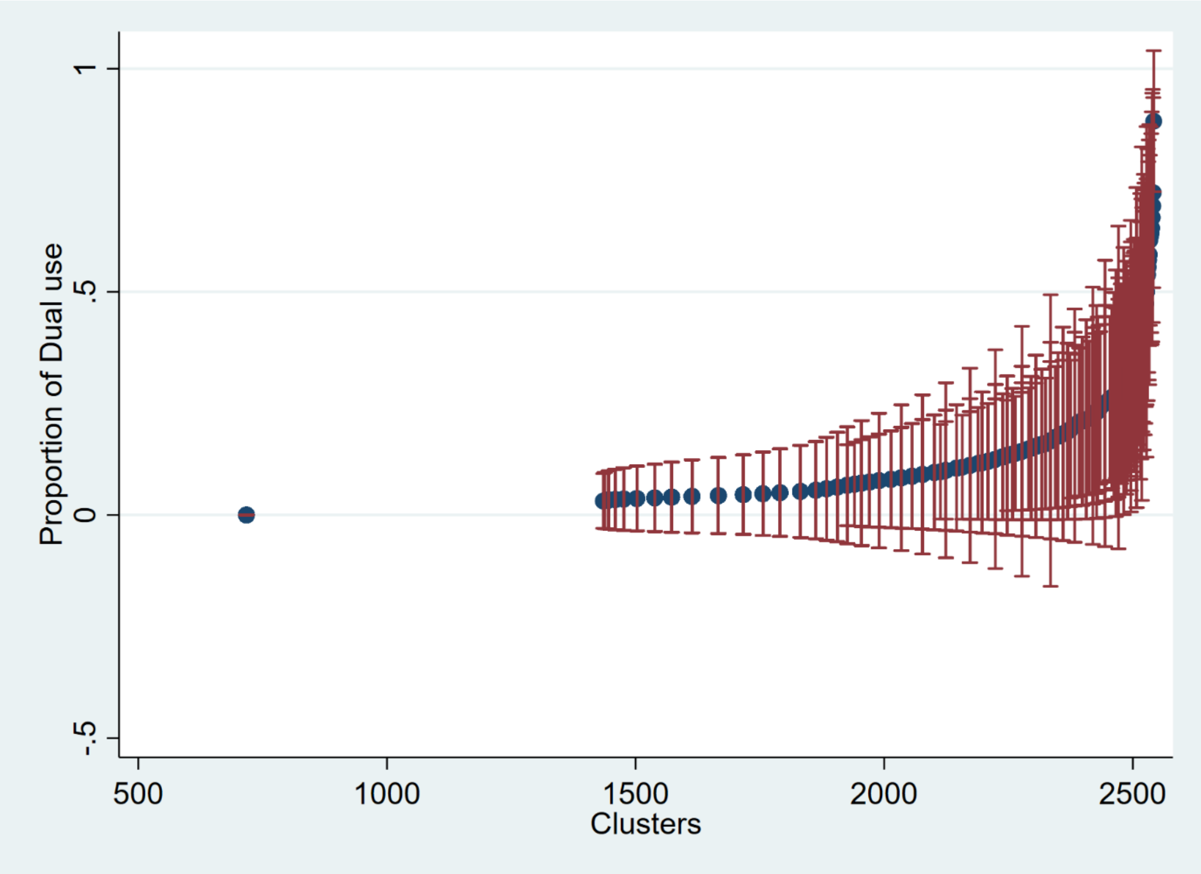
Prevalence and 95% confidence intervals of bidi smoking by local areas



Prevalence and 95% confidence intervals of SLT use by local areas



Prevalence and 95% confidence intervals of Dual use by local areas



Odds ratios for any tobacco use obtained from multilevel multivariable logistic regression models

Covariates	Categories	Odds Ratio	95% CI	
Age		1.02	1.02	1.03
Sex	Male			
	Female	0.23	0.21	0.24
Wealth	Poorer			
	Poor	0.85	0.81	0.90
	Middle	0.74	0.69	0.80
	Rich	0.58	0.53	0.63
	Richer	0.43	0.40	0.48
Education	No formal education			
	Less than primary	0.86	0.80	0.92
	Primary less than secondary	0.68	0.64	0.73
	Secondary and above	0.37	0.35	0.40
Occupation	Unemployed			
	Labourer	1.92	1.74	2.12
	Housewife/ Retired/ Student	0.66	0.60	0.73
	Self	1.54	1.40	1.70
	Private	1.53	1.37	1.72
	Government	1.14	1.00	1.30
Area of residence	Urban			
	Rural	1.11	1.04	1.19
States	Jammu & Kashmir			
	Himachal Pradesh	0.81	0.64	1.03
	Punjab	0.71	0.55	0.90
	Chandigarh	0.73	0.56	0.94
	Uttarakhand	1.44	1.15	1.81
	Haryana	1.20	0.95	1.52
	Delhi	1.18	0.92	1.51
	Rajasthan	1.06	0.85	1.32
	Uttar Pradesh	2.34	1.90	2.88
	Chhattisgarh	3.06	2.43	3.86
	Madhya Pradesh	1.78	1.44	2.21
	West Bengal	1.92	1.55	2.39
	Jharkhand	2.48	1.96	3.13
	Odisha	3.87	3.06	4.90
	Bihar	1.16	0.94	1.45
	Sikkim	0.86	0.66	1.12
	Arunachal Pradesh	3.74	2.90	4.82
	Nagaland	5.06	3.94	6.48
	Manipur	9.46	7.37	12.14
	Mizoram	4.64	3.61	5.95
	Tripura	11.87	9.21	15.29
	Meghalaya	3.06	2.38	3.92
	Assam	4.36	3.52	5.41
	Gujarat	1.26	1.01	1.58
	Maharashtra	1.39	1.12	1.72

	Goa	0.44	0.34	0.58
	Andhra Pradesh	0.50	0.39	0.63
	Telangana	0.68	0.53	0.87
	Karnataka	1.04	0.84	1.30
	Kerala	0.57	0.44	0.73
	Tamil Nadu	0.80	0.64	1.00
	Puducherry	0.53	0.41	0.68

Odds ratios for cigarette smoking obtained from multilevel multivariable logistic regression models

Covariates	Categories	Odds Ratio	95% CI	
Age		1.01	1.00	1.01
Sex	Male	0.03	0.02	0.04
	Female			
Wealth	Poorer			
	Poor	1.31	1.03	1.68
	Middle	1.80	1.38	2.35
	Rich	1.86	1.40	2.46
	Richer	1.80	1.36	2.40
Education	No formal education			
	Less than primary	1.14	0.86	1.51
	Primary less than secondary	1.24	0.98	1.56
	Secondary and above	1.02	0.81	1.30
Occupation	Unemployed			
	Labourer	1.70	1.24	2.34
	Housewife/ Retired/ Student	0.51	0.37	0.72
	Self	1.51	1.12	2.05
	Private	1.54	1.11	2.14
	Government	1.32	0.94	1.87
Area of residence	Urban			
	Rural	0.66	0.56	0.77
States	Jammu & Kashmir			
	Himachal Pradesh	0.13	0.08	0.21
	Punjab	0.08	0.04	0.13
	Chandigarh	0.10	0.06	0.16
	Uttarakhand	0.16	0.10	0.26
	Haryana	0.06	0.03	0.10
	Delhi	0.13	0.09	0.21
	Rajasthan	0.05	0.03	0.09
	Uttar Pradesh	0.05	0.03	0.10
	Chhattisgarh	0.03	0.01	0.09
	Madhya Pradesh	0.04	0.02	0.08
	West Bengal	0.19	0.12	0.30
	Jharkhand	0.21	0.12	0.37
	Odisha	0.09	0.04	0.19
	Bihar	0.02	0.01	0.06
	Sikkim	0.46	0.30	0.70
	Arunachal Pradesh	0.59	0.37	0.96
	Nagaland	0.07	0.03	0.16
	Manipur	1.05	0.69	1.60
	Mizoram	0.58	0.39	0.86
	Tripura	0.99	0.63	1.58
	Meghalaya	0.58	0.38	0.90
	Assam	0.22	0.13	0.35
	Gujarat	0.03	0.02	0.06

	Maharashtra	0.07	0.04	0.11
	Goa	0.06	0.03	0.11
	Andhra Pradesh	0.21	0.14	0.32
	Telangana	0.27	0.18	0.42
	Karnataka	0.16	0.10	0.24
	Kerala	0.26	0.17	0.39
	Tamil Nadu	0.27	0.19	0.38
	Puducherry	0.29	0.20	0.42

Odds ratios for bidi smoking obtained from multilevel multivariable logistic regression models

Covariates	Categories	Odds Ratio	95% CI	
Age		1.04	1.03	1.04
Sex	Male	0.04	0.04	0.05
	Female			
Wealth	Poorer			
	Poor	0.82	0.73	0.92
	Middle	0.67	0.57	0.78
	Rich	0.45	0.37	0.54
	Richer	0.28	0.23	0.35
Education	No formal education			
	Less than primary	0.79	0.69	0.91
	Primary less than secondary	0.48	0.43	0.55
	Secondary and above	0.18	0.15	0.21
Occupation	Unemployed			
	Labourer	2.45	2.02	2.96
	Housewife/ Retired/ Student	0.73	0.59	0.90
	Self	1.87	1.55	2.26
	Private	1.34	1.06	1.70
	Government	1.47	1.10	1.95
Area of residence	Urban			
	Rural	1.63	1.41	1.89
States	Jammu & Kashmir			
	Himachal Pradesh	3.24	2.19	4.78
	Punjab	1.18	0.77	1.82
	Chandigarh	2.49	1.59	3.90
	Uttarakhand	3.71	2.51	5.49
	Haryana	4.69	3.18	6.92
	Delhi	2.38	1.51	3.75
	Rajasthan	1.40	0.96	2.03
	Uttar Pradesh	1.32	0.90	1.94
	Chhattisgarh	0.53	0.32	0.86
	Madhya Pradesh	0.83	0.56	1.24
	West Bengal	1.98	1.36	2.89
	Jharkhand	0.27	0.15	0.50
	Odisha	0.37	0.22	0.63
	Bihar	0.32	0.21	0.50
	Sikkim	0.15	0.08	0.29
	Arunachal Pradesh	0.50	0.28	0.89
	Nagaland	1.26	0.77	2.07
	Manipur	0.26	0.13	0.53
	Mizoram	0.11	0.05	0.25
	Tripura	4.73	3.02	7.42
	Meghalaya	2.29	1.49	3.52
	Assam	0.83	0.54	1.26

	Gujarat	0.93	0.62	1.39
	Maharashtra	0.25	0.16	0.41
	Goa	0.27	0.14	0.50
	Andhra Pradesh	0.51	0.33	0.78
	Telangana	0.52	0.33	0.81
	Karnataka	0.61	0.40	0.91
	Kerala	0.42	0.25	0.69
	Tamil Nadu	0.62	0.41	0.93
	Puducherry	0.23	0.13	0.41

Odds ratios for smokeless tobacco use obtained from multilevel multivariable logistic regression models

Covariates	Categories	Odds Ratio	95% CI	
Age		1.02	1.02	1.02
Sex	Male			
	Female	0.44	0.41	0.46
Wealth	Poorer			
	Poor	0.88	0.83	0.94
	Middle	0.75	0.69	0.82
	Rich	0.56	0.50	0.62
	Richer	0.38	0.34	0.43
Education	No formal education			
	Less than primary	0.90	0.83	0.97
	Primary less than secondary	0.74	0.69	0.79
	Secondary and above	0.43	0.39	0.46
Occupation	Unemployed			
	Labourer	1.81	1.61	2.04
	Housewife/ Retired/ Student	0.67	0.60	0.76
	Self	1.46	1.30	1.64
	Private	1.56	1.36	1.79
	Government	0.99	0.84	1.17
Area of residence	Urban			
	Rural	1.09	1.01	1.19
States	Jammu & Kashmir			
	Himachal Pradesh	0.49	0.31	0.76
	Punjab	1.97	1.37	2.82
	Chandigarh	1.23	0.82	1.84
	Uttarakhand	2.35	1.66	3.34
	Haryana	1.39	0.95	2.03
	Delhi	2.64	1.83	3.82
	Rajasthan	3.03	2.19	4.19
	Uttar Pradesh	8.56	6.27	11.67
	Chhattisgarh	14.66	10.56	20.35
	Madhya Pradesh	7.32	5.34	10.04
	West Bengal	5.72	4.15	7.87
	Jharkhand	10.67	7.66	14.85
	Odisha	17.56	12.62	24.42
	Bihar	5.49	4.01	7.53
	Sikkim	2.21	1.51	3.23
	Arunachal Pradesh	11.81	8.29	16.82
	Nagaland	18.80	13.30	26.57
	Manipur	36.81	26.07	51.99
	Mizoram	17.94	12.69	25.36
	Tripura	40.00	28.21	56.70

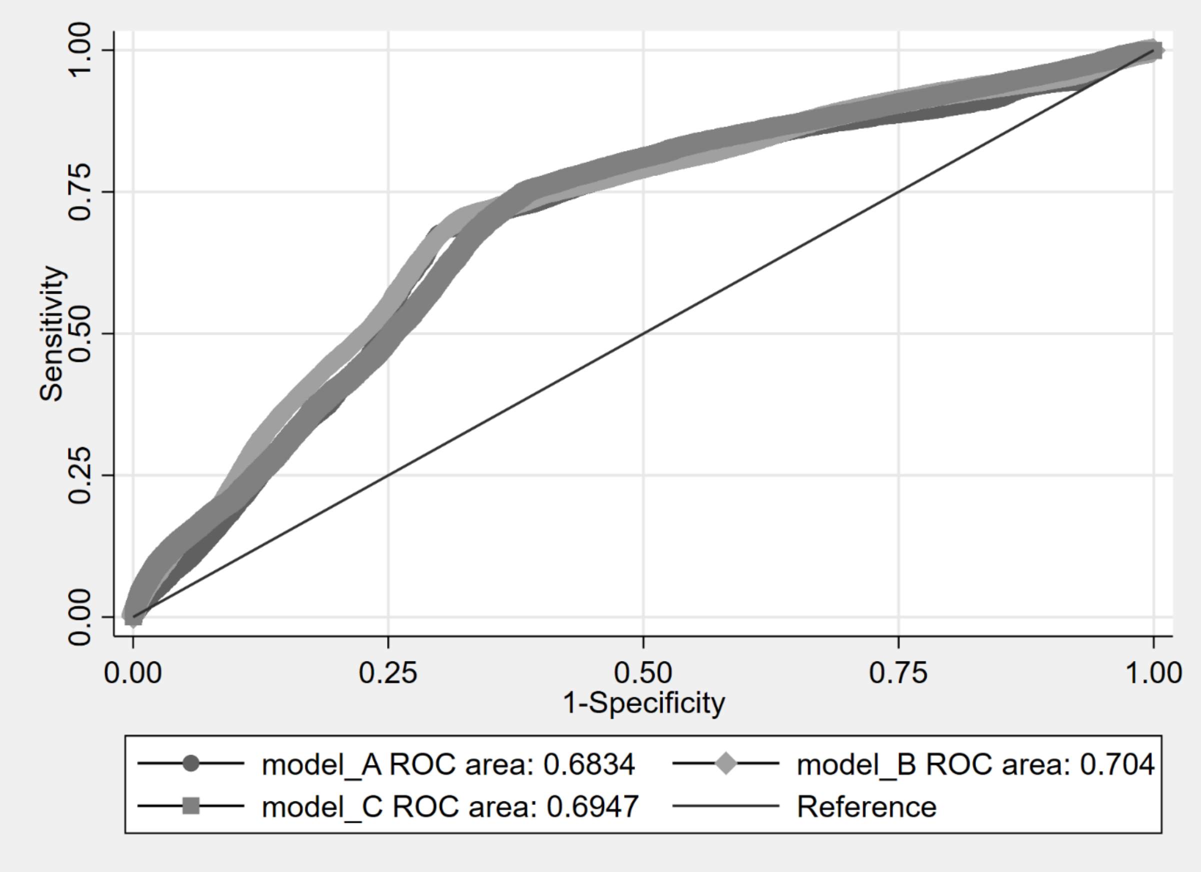
	Meghalaya	7.70	5.40	10.98
	Assam	17.68	12.92	24.20
	Gujarat	5.31	3.85	7.32
	Maharashtra	6.81	4.98	9.31
	Goa	1.76	1.21	2.56
	Andhra Pradesh	1.14	0.79	1.65
	Telangana	2.05	1.44	2.92
	Karnataka	3.63	2.62	5.02
	Kerala	1.12	0.76	1.65
	Tamil Nadu	2.15	1.54	3.00
	Puducherry	1.18	0.81	1.72

Odds ratios for dual use obtained from multilevel multivariable logistic regression models

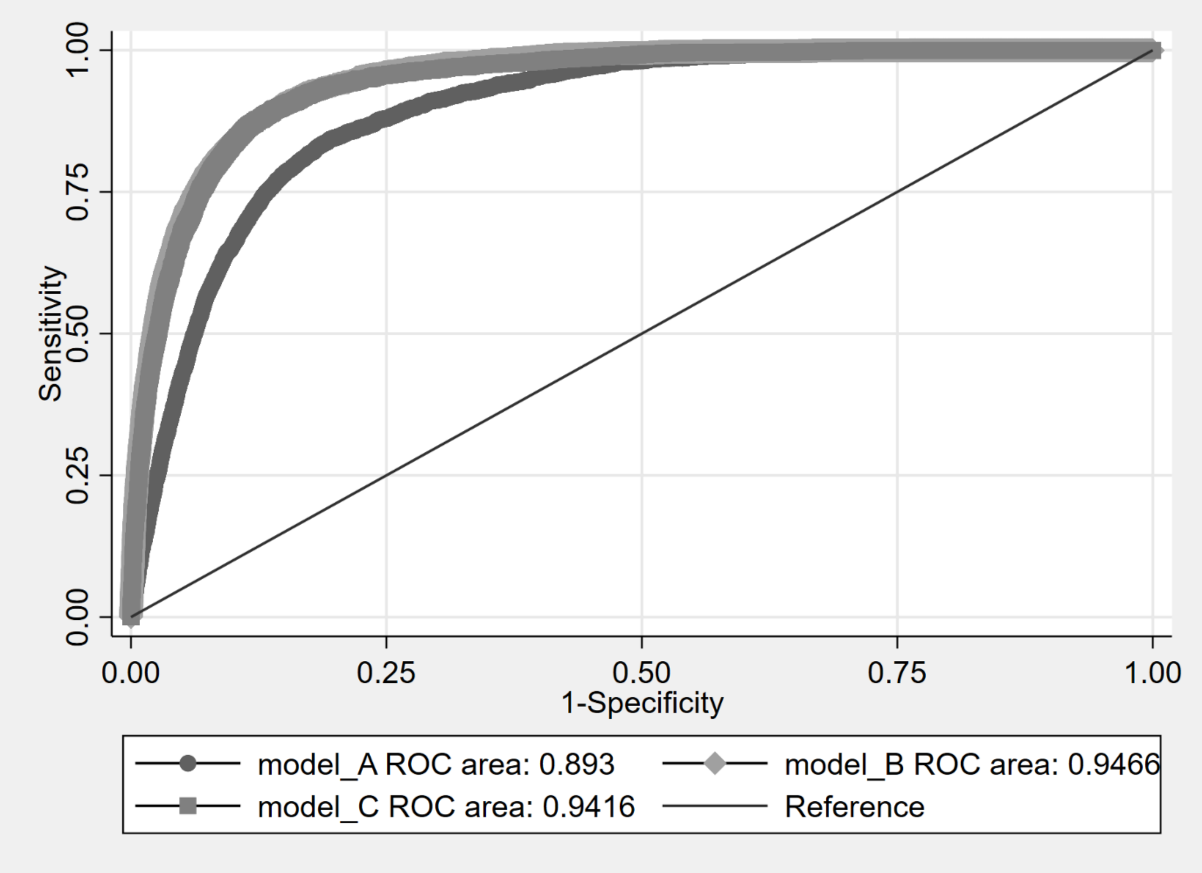
Covariates	Categories	Odds Ratio	95% CI	
Age		1.01	1.01	1.01
Sex	Male			
	Female	0.05	0.04	0.05
Wealth	Poorer			
	Poor	0.85	0.74	0.98
	Middle	0.74	0.61	0.89
	Rich	0.64	0.51	0.80
	Richer	0.45	0.35	0.58
Education	No formal education			
	Less than primary	0.82	0.69	0.98
	Primary less than secondary	0.62	0.54	0.73
	Secondary and above	0.26	0.22	0.31
Occupation	Unemployed			
	Labourer	2.18	1.72	2.77
	Housewife/ Retired/ Student	0.37	0.28	0.49
	Self	1.41	1.11	1.78
	Private	1.58	1.20	2.08
	Government	1.12	0.81	1.56
Area of residence	Urban			
	Rural	1.09	0.92	1.29
States	Jammu & Kashmir			
	Himachal Pradesh	1.00	0.49	2.04
	Punjab	1.37	0.70	2.68
	Chandigarh	1.34	0.65	2.74
	Uttarakhand	5.60	3.07	10.19
	Haryana	2.03	1.06	3.89
	Delhi	2.77	1.43	5.36
	Rajasthan	1.45	0.79	2.65
	Uttar Pradesh	7.29	4.17	12.74
	Chhattisgarh	2.06	1.06	3.99
	Madhya Pradesh	2.88	1.61	5.16
	West Bengal	3.06	1.70	5.52
	Jharkhand	7.41	4.04	13.57
	Odisha	4.55	2.44	8.48
	Bihar	1.73	0.95	3.14
	Sikkim	2.06	1.05	4.06
	Arunachal Pradesh	18.94	10.26	34.97
	Nagaland	17.87	9.68	33.00
	Manipur	28.55	15.51	52.53
	Mizoram	2.81	1.44	5.49
	Tripura	22.80	12.21	42.58
	Meghalaya	2.09	1.07	4.08
	Assam	7.16	4.04	12.69
	Gujarat	1.23	0.66	2.30

	Maharashtra	0.59	0.30	1.17
	Goa	0.30	0.11	0.81
	Andhra Pradesh	0.23	0.10	0.53
	Telangana	0.21	0.08	0.52
	Karnataka	1.33	0.72	2.46
	Kerala	1.19	0.59	2.37
	Tamil Nadu	0.65	0.34	1.27
	Puducherry	0.40	0.18	0.89

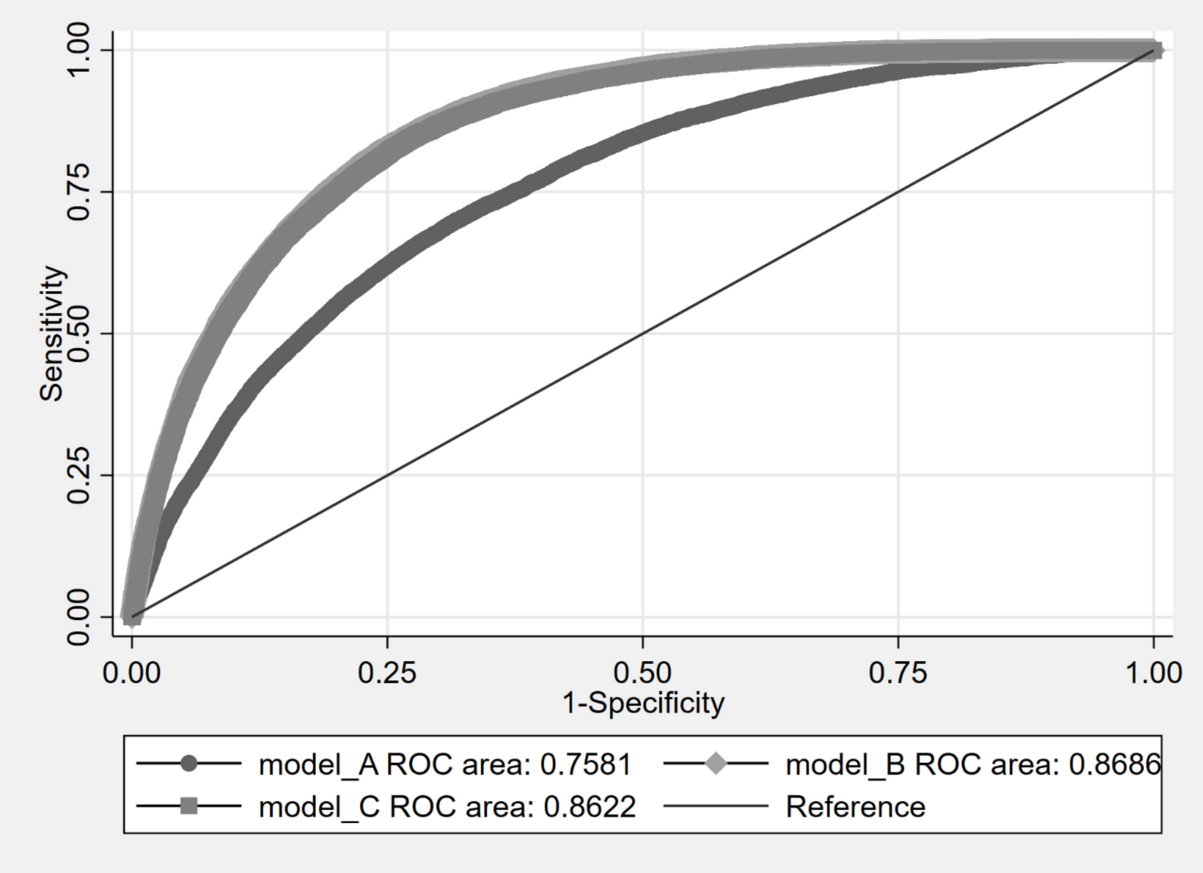
Area under the receiver operating characteristic (AU-ROC) curve for *cigarette smoking* plotted separately for single and multilevel logistic regression models



Area under the receiver operating characteristic (AU-ROC) curve for *bidi smoking* plotted separately for single and multilevel logistic regression models



Area under the receiver operating characteristic (AU-ROC) curve for *SLT use* plotted separately for single and multilevel logistic regression models



Area under the receiver operating characteristic (AU-ROC) curve for *dual use* plotted separately for single and multilevel logistic regression models

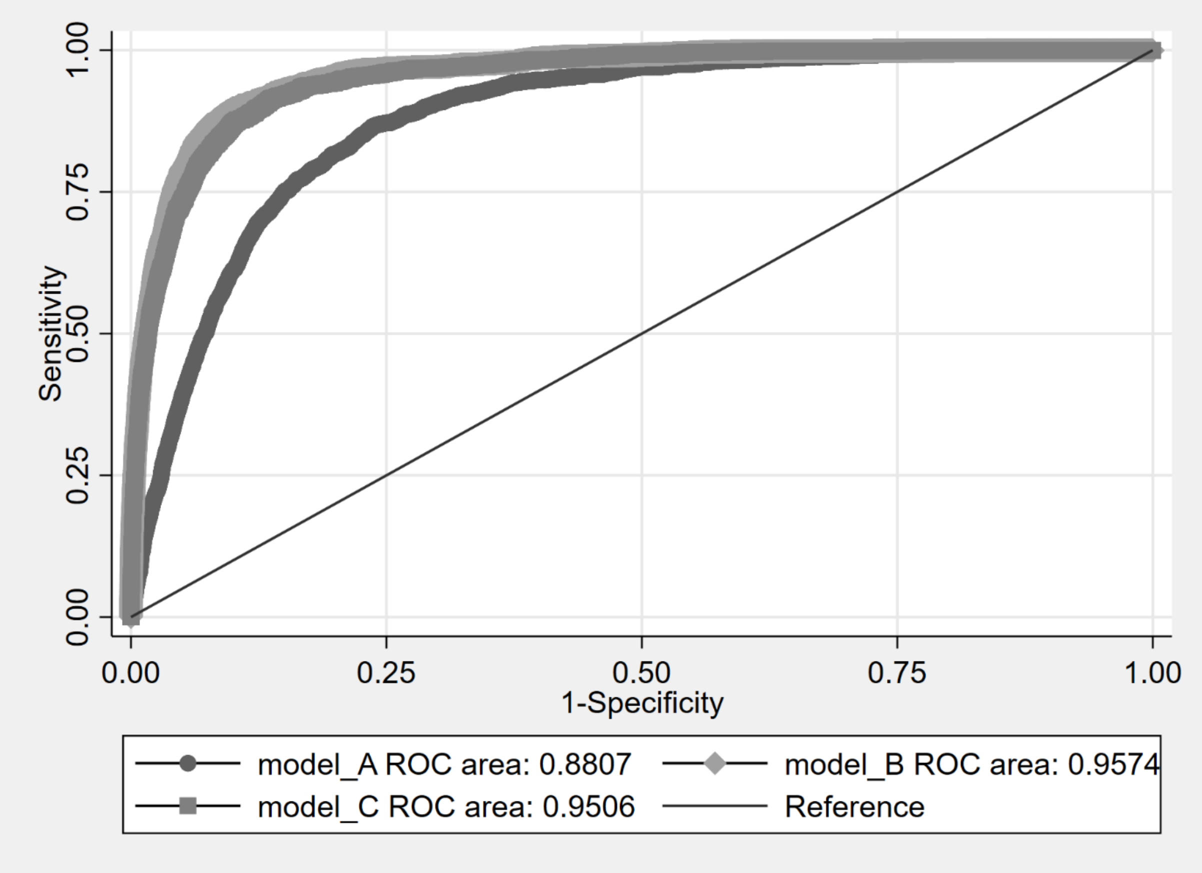


Table showing intraclass correlation coefficients obtained from three level hierarchical models with individual nested within city-wards/villages nested within states

		Null Model ICC 95% CI	Full model ICC 95% CI
Any tobacco use	State	17% (11%, 25%)	17% (11%, 25%)
	City ward/ village within state	23% (17%, 30%)	23% (17%, 30%)
Cigarette smoking	State	23% (15%, 34%)	22% (14%, 32%)
	City ward/ village within state	33% (25%, 42%)	33% (26%, 42%)
Bidi smoking	State	16% (11%, 25%)	19% (13%, 29%)
	City ward/ village within state	30% (24%, 36%)	33% (26%, 40%)
SLT use	State	28% (20%, 40%)	26% (18%, 36%)
	City ward/ village within state	36% (28%, 46%)	33% (25%, 42%)
Dual use	State	27% (18%, 38%)	26% (18%, 37%)
	City ward/ village within state	39% (31%, 48%)	43% (36%, 51%)

ICC: Intra-class correlation
Null model: Intercept only
Full model: Age, sex, education, occupation, wealth and area of residence

Interpretation

Example any tobacco use: Conditioned on covariates (age, sex, education, occupation, wealth and area of residence) any tobacco use is highly correlated within states (ICC: 17%). Within the same cityward/ village and state, this correlation was even higher (ICC: 23%).

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2-3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	5-6
Objectives	3	State specific objectives, including any prespecified hypotheses	6
Methods			
Study design	4	Present key elements of study design early in the paper	6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	7
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	7
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	7-8
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	7-8
Bias	9	Describe any efforts to address potential sources of bias	8-9
Study size	10	Explain how the study size was arrived at	7
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	8-9
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	8-10
		(*b*) Describe any methods used to examine subgroups and interactions	Not applicable
		(*c*) Explain how missing data were addressed	7
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	7

(*e*) Describe any sensitivity analyses	10

Continued on next page

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	9
		(b) Give reasons for non-participation at each stage	10
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	10-11
		(b) Indicate number of participants with missing data for each variable of interest	
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	10-11
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	11
		(*b*) Report category boundaries when continuous variables were categorized	Not applicable
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Not applicable
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	11
Discussion			
Key results	18	Summarise key results with reference to study objectives	10
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	10-11
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	11-12
Generalisability	21	Discuss the generalisability (external validity) of the study results	12
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	17

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.