Demography India (2018) ISSN: 0970-454X
Vol. 47, Issue: 1, pp: 61-70

Research Article

Epidemiological Modelling of Tobacco Use to Find Its Correlates among Adolescents in India: Comparative Appraisal of Conventional and Hierarchical Models

Ashish Datt Upadhyay[1], Sada Nand Dwivedi[1,*], Vishnubhatla Sreenivas[1] & Anju Dhawan[2]

Abstract

Tobacco use is the known risk factor of some of the major causes of death like cancer, respiratory diseases and heart diseases. In case of hierarchical structure of data, to obtain more accurate analytical results, multilevel regression model is often preferred in epidemiological and public health research. Aim of this study was to find factors associated with tobacco use among adolescents and appraisal of comparative results under conventional and multilevel logistic models. For this, available data on 37033 adolescents (age 15–19) years under National Family Health Survey-3 (2005–2006) were used. The performance of the models was assessed by the log likelihood values, area under the ROC curve, AIC and BIC. In this study, socio-demographic variables which are likely to be positively associated with the chance of tobacco use among adolescents were low education, employment, caste and religion other than Non-SC/ST, low wealth index, advancing age and male gender. Adolescents who belong to tobacco using household were more likely to use tobacco than their counterparts. Among the state level variables, adolescents belonging to states where prevalence of education of 10th standard and above was below 30.8%, or prevalence of tobacco users was above 34.6%**,** were more likely to use tobacco than belonging to otherwise. This study revealed that low education of adolescents and higher prevalence of tobacco use in the state are important factors to encourage adolescents regarding tobacco use. On other hand, although associated factors under multilevel model also remained similar, the model performance parameters showed efficacy of multilevel model in comparison to conventional logistic model.

Introduction

Tobacco use is the known risk factor of some of the major causes of death like cancer, respiratory diseases and heart diseases. It is second major cause of death in the world today (W.H.O Report on the Global Tobacco Epidemics, 2011). According to WHO estimate, during 20th century tobacco use killed hundred million of people and nearly each year 4.9 million deaths were attributed to tobacco use and which may rise to 10 million annually by 2020. Among these deaths, majority of tobacco attributed deaths come from developing countries like India.

According to NFHS-3, prevalence of tobacco use and smoking cigarettes or beedis among females of age group 15-19 years were 3.5% and 0.1% and among males were 28.6% and 12.3% respectively in the year 2005-2006. According to Global Youth Tobacco Survey (G.Y.T.S.), 2009, on youth of age 13 to 15 years, supported by World Health Organization (W.H.O.) and centre for Disease control and prevention, the prevalence of any type of tobacco use was 14% (Boys: 19%, Girls: 8.3%).

Adolescents and young adults are the future of any country. Without thinking about pros and cons, adolescents intend to try new things to taste the world and that is why they are always targeted by the tobacco companies. In India, 40.7% of the population account for less than 20 years of age and

[1]Department of Biostatistics, All India Institute of Medical Sciences (AIIMS), Ansari Nagar, New Delhi-29
[2]National Drug Dependence Treatment Center, Department of Psychiatry, AIIMS, Ansari Nagar, New Delhi-29
*Corresponding Author: Prof. Sada Nand Dwivedi, Email: dwivedi7@hotmail.com

20.9% falls in age group 10-19 years (Census-2011). Therefore, this age group needs to be handled carefully because they are more likely to adapt the risk behaviours like tobacco use. Also, in our country, very little attention has so far been focused on understanding the determinants of tobacco use among adolescents.

In areas of the social, medical and other sciences, characteristics of individuals get influenced by neighbourhood like community, district or state where they reside. In case of binary outcome, conventional logistic regression does not involve prevailing hierarchical structure of data and assumes that all records are independent {See Raudenbush and Bryk, 1992}. Hence, due to distortion in required assumption of independence, standard error of the estimated parameter gets underestimated. As a result, inaccurate analytical results are obtained leading to inappropriate public health planning. Therefore, aim of this study was to find out factors associated with current tobacco use among adolescents and their comparison under the two statistical modelling approaches.

Material and Methods

For this study, data on adolescents of age group 15-19 years was extracted from third round of the National Family Health survey (NFHS-3). This survey was done across India among 29 states in 2005–06. NFHS-3 mainly provides national and state level estimates of fertility, family planning, infant and child mortality, reproductive and child health, nutrition of women and children, and the quality of health and family welfare services. In NFHS-3, a total of 124,385 women aged 15–49 years and 74369 men aged 15–54 years were interviewed from 109041 households. These household were selected using multistage sampling design with two stage design in rural areas and a three stage design in urban areas. The detail methodology is available in its national level report {IIPS and Macro International. 2007, NFHS-3, 2005–06: India: Volume I &Volume II}.

In the survey questionnaires, there were three questions addressing about self reported current tobacco use. They were “Do you currently smoke cigarettes or bidis?”; “Do you currently smoke or use tobacco in any other form?”; and “In what other form do you currently smoke and/or use tobacco?”. The current tobacco use defined for this study was “Use of smoke and/or smokeless tobacco use”.

On the basis of review of literatures and subject knowledge, a set of independent or exploratory variables were selected for the analysis. After exploratory analysis on original forms of variables in original data, some of them were retained in their existing forms and others were modified to get meaningful and stable estimates. The selected independent/ exploratory variables considered in their existing forms were: age; sex (male/female); place of residence (rural /urban); employment (yes/no); house hold structure (nuclear/non-nuclear); employment of adolescent (yes/no); and alcohol use of adolescent (yes/no). Further, to enable meaningful analysis, some qualitative variables were considered after appropriate modification: adolescent’s education (below secondary/ secondary and above); wealth index of household (poor or poorest/middle/high or highest), and adolescent’s marital status (married/un-married or single). Also, as considered by Pandey et al. (1998), religion and caste were pooled to derive another variable religion/caste (SC-ST Hindu/other Hindu/Muslim /other religion).

Adolescent’s exposure to family tobacco use (yes /no) was derived from tobacco use information of all surveyed men and women in NFHS-3. Adolescents were categorised as exposed to household tobacco use, if any other family member was using tobacco in household. Adolescent exposure to media was derived by information available in NFHS-3 like they listen radio, read news paper/ magazine, watch television and watch cinema monthly. The adolescents were categorized as exposed to media, if adolescent read paper / listen radio/ watch television / watch cinema monthly. Some state level variables such as state literacy $10^{th}$and above standard among 15 years and above year of age ($<30.8\%$or.$\geq 30.8\%$) were included. Among these variables, state literacy prevalence was extracted from census 2011 data, and state tobacco prevalence was taken from Global Adult Tobacco survey (GATS) report 2009-10. The threshold level for each variable was chosen as the national prevalence of the respective variables.

Ethics Statement

Survey data available in public domain for academic use (http://www.measuredhs.com) was utilized and it did not require ethics committee approval.

Statistical Models

Logistic regression describes the relationship between binary outcomes like current tobacco use among adolescents with other associated variables. The logistic model describes the probability p_{ij} that the i^{th} adolescent in the j^{th} state is likely to use tobacco of as a function of the considered associated variables as (Goldstein, H. (2011)):

$$\mathrm{Log}_e\left[\frac{p_{ij}}{1-p_{ij}}\right]=\beta_0+\sum_{K=1}^{m}\beta_K X_{ijk}$$

Where, β_0 is a constant; β_k is the regression coefficient of the k^{th} associated factor and X_{ijk} represents i^{th} adolescent's value in j^{th} state for k^{th} associated factor. As true in present case, when both state level and individual level characteristics are present in data, conventional logistic regression disaggregates the state level variables at the individual level and treats as adolescent's level variables which obviously distorts the assumptions and may influence standard error estimates.

In India, different states have varying traditions and socio demographic environments. To accommodate this neighbourhood effect, multivariable multilevel logistic regression with random intercept analysis was used. Multilevel model is the extension of logistic regression model. Logistic model becomes multilevel model if it includes the state level random effect (u_i) as predictor. State level random effect is a totality of measured and unmeasured state-level variables that predict tobacco use and are uncorrelated with the individual and state level associated variables in the model. When data have only two levels like individual and state level and p^*_{ij} is the conditional probability that adolescent i in state j used tobacco, two level multilevel random intercept model is defined as {Goldstein, H. (2011)}:

$$\mathrm{Log}_e\left[\frac{p^*_{ij}}{1-p^*_{ij}}\right]=\mu_i+\beta_0+\sum_{K=1}^{m}\beta_K X_{ijk}$$

In beginning, to assess the association of each factor with adolescent tobacco use, association analysis for each exploratory variable was carried out against the dependent variable using chi-square test. Stepwise logistic regression was adopted for conventional multivariable regression technique. All the individual and state level variables found significant at 25% level of significance in association analysis were considered for stepwise multivariable logistic regression model. An entry probability of 0.1 and an exit probability of 0.15 were used for the stepwise model. The maximum likelihood approach was used for parameter estimation.

During model building process, possible multi-co linearity and first order interaction effect were investigated. Non-occurrence of multi-co linearity was assessed for covariate with cut-off point of mean of variation inflation factor (VIF) as less than five (Garson,2012).In other words, if the model mean VIF was greater than five then variables causing multi-co linearity were identified and removed. Effect modifier was assessed by two approaches; one was stratified analysis with confidence interval approach and second was multivariable regression approach. No variable was found as effect modifier in this study.

Keeping in view of comparative appraisal, multivariable multilevel logistic regression with random intercept analysis was also conducted on same set of covariates as those under multivariable logistic regression. The log likelihood for this model was approximated by maximum likelihood estimation with adaptive Gaussian quadrature {Pinheiro, J. C., (2006)}

` The area level variance has been assessed by median odds ratio, 80% interval odds ratio, and intra-class correlation. The social, economic, health facilities and other characteristics of one area are likely to be different from another. Further, all adolescent in the same area share a common set of environment. Intra-class correlation measures the proportion of total variance in the outcome that is attributable to the area level characteristics {See Larsen & Merlo (2005)}. Median Odds Ratio (MOR) proposed by Larsen & Merlo (2005) transforms the area level variance to the odds ratio scale. It is the median of the set of odds ratios that could be obtained by comparing the two randomly chosen adolescent's one from high risk area and other from low risk area with identical individual level covariates. If the MOR is 1, the area level variation is close to zero and >1, there exits substantial cluster level variation. Another index used for explaining the area level variability is the 80% Interval Odds Ratio (IOR-80) {See, Larsen & Merlo (2005)}. It is a fixed-effects measure for quantification of the effect of cluster-level variables. IOR-80 for a given state level variable is the middle 80% of the range of all odds ratio calculated from each pair of all possible pairs of adolescents with identical individual-level risk factors from different state but who differ by one level in state-level risk factors. If the IOR-80 is narrower and does not contain unity means specific state level variable explains the area level variability substantially. However, if interval is wider or contains 1, it implies that specific state level variable explains the area level variability minimally {See, Larsen & Merlo (2005)}.

To compare the predictive performance of the models, various indices such as Log-likelihood, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and area under receiving operative curve analysis (ROC) were obtained. Lower values of AIC and BIC, and higher values of Log-likelihood and ROC indicate better fit of the model. The analysis was carried out on STATA software (version 14) and p-value less than 0.05 was considered as significant.

Results

In NFHS-3, out of 37033 adolescents (15-19 years), 4802 (12.97%) were current tobacco users (smoke or smokeless tobacco user). Mean age of adolescents was 17.0±1.38 years. Almost sixty five percent of surveyed adolescents were female and majority of them from rural area (54%). Bivariate analysis showed (Table- 1a&b) that adolescents who were male (OR=7.0, 95%CI: 6.53–7.50), belong to non nuclear family (OR=1.2, 95%CI: 1.15–1.29) and educated below secondary school (OR=1.9, 95%CI: 1.77-2.01) were significantly more likely to be current tobacco users. On other-hand, tobacco use significantly increased with decrease in wealth status, and 10th standard and above state level literacy.

Table 1-a. Association of current tobacco use among adolescents with socio-economic & demographic factors

Characteristic	Current Tobacco Use Yes f(%)	No f(%)	Unadjusted Odds Ratio (95%C.I.)
Age(years) mean±sd	17.4±1.3	16.9±1.4	1.2 (1.24,1.31)
Sex			
Female	1223 (5.1)	22730 (94.9)	1.0
Male	3579 (27.4)	9499 (72.6)	7.0 (6.53,7.50)
Residence			
Urban	2000 (11.7)	15023 (88.3)	1.0
Rural	2802 (14.0)	17206 (86.0)	1.2 (1.15,1.30)
Household structure			
Nuclear	2371 (11.9)	17491 (88.1)	1.0
Non-nuclear	2431 (14.2)	14738 (85.8)	1.2 (1.14,1.29)
Wealth Index			
Richer/Richest	1925 (9.8)	17722 (90.2)	1.0
Middle	1157 (14.9)	6630 (85.1)	1.6 (1.48,1.73)
Poorest/Poorer	1720 (17.9)	7877 (82.1)	2.0 (1.87,2.16)
Caste or Religion			

Hindu(non SC/ST)	1692 (9.4)	16244 (90.6)	1.0
Hindu(SC/ST)	1232 (15.8)	6582 (84.2)	1.8 (1.66,1.94)
Muslim	728 (12.4)	5119 (87.6)	1.4 (1.24,1.49)
Others religions	1125 (21.6)	4090 (78.4)	2.6 (2.40,2.87)
Education			
Secondary & above	3023 (10.9)	24572 (89.1)	1.0
Illiterate/Primary	1777 (18.8)	7650 (81.2)	1.9(1.77, 2.01)
Marital status			
Unmarried/Single	4227 (13.2)	27737 (86.8)	1.0
Married	575 (11.4)	4492 (88.6)	0.8 (0.77,0.92)
Employment			
Unemployed	2048 (8.1)	23322 (91.9)	1.0
Employed	2754 (23.8)	8828 (76.2)	3.6 (3.34,3.78)
Exposure to media			
No	305 (9.8)	2801 (90.2)	1.0
Yes	4497 (13.3)	29410 (86.7)	1.4 (1.24,1.58)
Alcohol use			
No	3507 (9.9)	31634 (90.1)	1.0
Yes	1292 (68.6)	592 (31.4)	19.7 (17.75,21.83)
Employment			
Unemployed	2048 (8.1)	23322 (91.9)	1.0
Employed	2754 (23.8)	8828 (76.2)	1.2 (1.24,1.31)
Media exposed			
No	305 (9.8)	2801 (90.2)	1.0
Yes	4497 (13.3)	29410 (86.7)	1.4 (1.24,1.59)

Table 1-b: Association of current tobacco use with household level& states level factors

Characteristic	**Current Tobacco Use** Yes f(%)	No f(%)	**Un-adjusted Odds Ratio (95%C.I.)**
Family tobacco use			
No	1107 (6.7)	15377 (93.3)	1.0
Yes	3295 (18.6)	14453 (81.4)	3.2 (2.94,3.40)
Literacy rate (10 and above) in State			
≥30.8%	1427 (8.5)	15442 (91.5)	1.0
<30.8%	3375 (16.7)	16787 (83.3)	2.2 (2.04,2.32)
Prevalence of tobacco use state(GATS data 2009-2010)			
≤ 34.6%	2063 (9.0)	20831 (91.0)	1.0
> 34.6%	2739 (19.4)	11398 (80.6)	2.4 (2.28,2.58)

At multivariable logistic regression level, the statistically significant socio- demographic, household and state level variables associated with tobacco use were (shown in table-2 a & b) gender, age, caste or religion, wealth status, marital status, type of family, employment, education, alcohol use of adolescents, household member tobacco use, adult tobacco use prevalence and state level 10th and above education proportion. Adolescents who were male (OR= 7.2, 95% CI: 6.57–7.94), belong to low wealth status (OR=1.4, 95% CI:1.25–1.53) or middle wealth status (OR=1.3, 95% CI: 1.22–1.48) as compared to richer/richest wealth status, belong to Schedule tribe or caste (OR=1.4, 95%CI: 1.24–1.52) or Muslim (OR=1.5, 95% CI: 1.34–1.68) or other religion (OR= 2.1, 95% CI: 1.87–2.34) as compared to Non–Scheduled cast or tribe and married (OR=1.3, 95% CI: 1.10–1.47) were significantly more likely to use tobacco. In terms of education and employment, adolescents who had education below secondary (OR=1.6, 95% CI: 1.48–1.79) and who were employed (OR=1.9, 95% CI:

1.83–2.16) were more likely to use tobacco. However, alcohol using adolescents had a higher chance (OR=8.3, 95% CI: 7.31–9.39) of using tobacco as compared to non alcohol users. In term of household and state level variables, adolescents residing in household whose other members were using tobacco (OR=2.2, 95% CI: 2.02–2.39), belonging to state where education level of 10th and above is less than 30.8% (OR=1.7, 95% CI: 1.52–1.84) and higher prevalence of tobacco use (>34.6%) in youth (OR=2.2, 95% CI: 2.02–2.44) were significantly more likely to use tobacco.

Multilevel analysis revealed that there were variation in tobacco use among the states and proportion of the estimated variance in tobacco use among adolescents between states was 10% (ICC=8%, 95% CI: 6.1%-16.5%). In term of odds, it is represented in Median odds ratio (MOR) and in this study, if adolescents moved to another state with a higher probability of adolescents tobacco use, the median increase in their odds of tobacco use would be 1.8-fold (MOR=1.8, 95% CI: 1.55-2.16). Table 2b reveals that nearly all the significantly associated factors in logistic regression remained also significant in multilevel analysis. However, their confidence interval becomes wider in multilevel model. It may be attributed to fact that conventional logistic regression does not account the state heterogeneity. Secondly, in case of multilevel logistic regression, odds of tobacco use by adolescents having education below secondary (OR=1.7, 95% CI: 1.58–1.93) is interpreted as if we compare two adolescents with identical level of associated factors, one with below secondary education and one with secondary and above, limited to the same state, then the chance of tobacco use was increased by 1.7 times for the adolescents having below secondary education.

In case of state level variables, this interpretation is limited to state having same level of random effect or risk level of tobacco use. In multilevel model, the odds of tobacco use for adolescents for state with lower literacy (<30.8%) of 10th and above as compared to state with higher literacy of 10th and above standard (≥30.8%) was 1.9 (95% CI: 0.95-3.64), limited to the states with the same risk of tobacco use. In other words, assuming otherwise comparable risk of tobacco use, chance of tobacco use by adolescents residing in the state having lower literacy of 10th standard and above (<30.8%) was 1.9 (95% CI: 0.95–3.64) times as compared to their counterparts.

Interval odds ratio 80% (IOR-80%) was calculated for quantifying fixed-effects measure of state level variables like state 10th and above literacy rate and adult tobacco prevalence. The IOR-80% for state education prevalence of 10th and above (<30.8%) was 0.62 to 4.28, and it provides the insight that when compared two randomly chosen adolescents with identical associated factors except adolescents one belonging to state with lower prevalence of literacy of 10th and above (< 31%), and other from its counterpart and those states possibly differ in tobacco use risks, the odds ratio for the comparison will lie between 0.62 to 4.28 with 80% probability. In case of state adult tobacco use prevalence (>34.6), IOR-80% was (0.82, 7.59). Since IOR-80% for state (cluster) variables were wider and include 1, it shows inability of state level factors to add meaningfully explanation of variation in prevalence of tobacco use in adolescents among the states.

Table 2-a: Multivariable association of individual study variables with current tobacco use

Characteristic	Logistic Regression Adjusted Odds Ratio (95%C.I.)	Multilevel Logistic Regression Adjusted Odds Ratio (95%C.I.)
Age (years)	1.3 (1.23,1.31)	1.3 (1.23,1.31)
Sex		
Female	1.0	1.0
Male	7.2 (6.57,7.94)	7.4 (6.69,8.12)
Household structure		
Nuclear	1.0	1.0
Non-nuclear	1.1 (1.03,1.20)	1.1 (1.05,1.23)
Wealth Index		
Richer/Richest	1.0	1.0
Middle	1.3 (1.22,1.48)	1.3 (1.16,1.43)
Poorest/Poorer	1.4 (1.25,1.53)	1.5 (1.32,1.64)
Caste or Religion		

Hindu(non SC/ST)	1.0	1.0
Hindu(SC/ST)	1.4 (1.24,1.52)	1.4 (1.24,1.52)
Muslim	1.5 (1.34,1.68)	1.4 (1.28,1.63)
Others religions	2.1 (1.87,2.34)	1.3 (1.16,1.57)
Education		
Secondary & above	1.0	1.0
Illiterate/Primary	1.6 (1.48,1.79)	1.7 (1.58,1.93)
Marital status		
Unmarried/Single	1.0	1.0
Married	1.3 (1.10,1.47)	1.4 (1.19,1.59)
Employment		
Unemployed	1.0	1.0
Employed	1.9 (1.83,2.16)	2.0 (1.85,2.20)
Exposure to media		
No	1.0	1.0
Yes	1.7 (1.40,1.99)	1.6 (1.33,1.90)
Alcohol use		
No	1.0	1.0
Yes	8.3 (7.31,9.39)	10.6(9.25,12.12)

Table 2-b: Multivariable association of individual study variables with current tobacco use and measures of district level variation

Characteristic	Logistic Regression Adjusted Odds Ratio (95%C.I.)	Multilevel Logistic Regression Adjusted Odds Ratio (95%C.I.)
Household tobacco use		
No	1.0	1.0
Yes	2.2 (2.02,2.39)	1.9 (1.76,2.09)
Literacy rate (10 and above) in State		
≥30.8%	1.0	1.0
<30.8%	1.7 (1.52, 1.84)	1.9 (0.95,3.64)
Interval (80%) odds ratio		{0.62,4.28)
Prevalence of tobacco use in state(GATS009-10)		
≤ 34.6%	1.0	1.0
>34.6%	2.2 (2.02,2.44)	2.5 (1.28,4.86)
Interval (80%) odds ratio		{0.82,7.59}
Measures of state level variation		
State level variance (95% C.I)		
Full Model	-----	0.37 (0.21, 0.65)
ICC(State level) {95% C.I}		
Full model	-------	0.1 (0 .06,0.16)
Median odds ratio (95% C.I.)		
Full Model	--------	1.8 (1.55,2.16)

Further, discriminating ability of two modelling approach shows that area under Receiving Operative Curve analysis (ROC) was more under multilevel model (AROC= 88.8, 95% CI: 88.3-89.2) as compared to conventional logistic model (AROC=87.5, 95% CI: 86.99-88.01). Table 3c shows

that multilevel model had largest log likelihood and the smallest AIC and BIC, as compared to standard logistic regression, suggesting best goodness of fit.

Table 3a. Comparison of models Measures of for current tobacco use among adolescents in India

	Conventional Logistic Regression	Multilevel Logistic Regression
Log likelihood	-8933.45	-8670.15
AIC	17900.9	17376.3
BIC	18044.24	17528.08
Area under ROC	87.5 (86.99,88.01)	88.8 (88.3,89.2)

Discussion

Tobacco use in adolescents has many grisly impacts on their future life and this study investigated associated factors related to the adolescents' tendency towards tobacco use and also demonstrates the comparison of models in terms of their predicting ability about adolescents current tobacco use. Multivariable analysis showed that several factors were associated with current tobacco use among adolescents. Among them, important factors which increase the chance of tobacco use were house-hold tobacco use, and state having higher tobacco use prevalence or environment. Similar finding was reported in other studies (Pinilla et al.2002; Ravishankar et al., 2009; Awasthi et al. 2010) and it confirms the social learning in initiation of tobacco use. It is evident from this study that community interventions may be required that include supportive environment, strong policy support and community participation. It is highly imperative to educate parents and society that their children are more likely to use tobacco whose family member use tobacco. Male and older adolescents had higher odds of using tobacco and this finding was also seen in other studies (Sreeramareddy et al., 2008; Rice et al., 2006). However, marital status variable changes from protective factor in univariable to risk factor in multivariable regression analysis. This change is likely to occur in study because in population, nearly 65% of adolescents were women and among them 20% were married, where as among males (35%) only 2 % percent were married. In multivariable analysis, average level of other variables like sex is considered, so the risks direction may change.

The present study has showed that literacy plays an important role in protecting adolescents from using tobacco. The adolescents having lower literacy and belonging to state where literacy was low were more likely to use tobacco than having higher literacy or belonging to state having higher literacy. Other studies have also reported similar findings (Rani et al., 2003; Sorensen et al., 2005; Subramanian et al., 2004; Mathuret.al 2008}. The reason behind this may be that due to literacy people have better understanding about the health risks associated with tobacco use. Alcohol use was a major correlate of tobacco use in our study. Combined use of alcohol and tobacco has been reported in other parts of the world (Jackson et al., 2003; Yawson et al., 2013) and in India (Gupta et al., 2005). Tobacco use was more likely to exist among Hindus (S.C./S.T.) and Muslims, compared to Non-Hindu (S.C/S.T.). This finding is analogous with other study (Subramanian et al., 2004). Lower socioeconomic status and prevalent low literacy among these groups may be major reasons behind it.

In terms of performance of models, discriminating ability of multilevel model was higher as compared to conventional logistic model. AIC and BIC also show that multilevel model is better than conventional logistic model. But, 95% confidence interval of significant factors was wider than standard logistic model (Sanagou et al., 2013). The expected reason may be that conventional logistic regression does not account state heterogeneity. Further, MOR suggests that there was variation in tobacco use among the states but 80% IOR conveys that state level variables did not give much information regarding the variation in tobacco use in adolescents among state. In other words, other cluster-level variables may be needed to explain the cluster heterogeneity. The main limitation of this data is that major focus of NFHS-3 survey was not on tobacco use hence information about many important variables could not be available.

Conclusion

This study has dealt with the application, interpretation and comparison of conventional and multilevel logistic regression regarding determinants of tobacco use among the adolescents. To best of our knowledge, our study is first study on such a large national level data to examine the associated factors of tobacco use among adolescents. In our study, multilevel model outperformed conventional model due to obvious presence of clustering in our data. This study has showed that education level of adolescents and community and tobacco use in their family and community were main predictors of tobacco use among adolescents. Keeping in view of associated factors, the preventive activity to curb tobacco use among them can be done by encouraging them regarding higher education level and also changing the social norm of tobacco use among the parents and society at home as well as at public places. Since data was not collected solely for tobacco use, further researches are needed to explore the vulnerability of certain more variables associated with tobacco use to generate effective programs.

Acknowledgement

The authors acknowledge the All India Institute of Medical Science, New Delhi, for providing required resources and financial support during study.

Reference

Awasthi, S., Jha, S. K., Rawat, C. M. S., Pandey, S., & Swami, S. S. (2010). Correlates of tobacco use among male adolescents in schools of Haldwani, Nainital. *Health and Population: Perspectives and Issues*, *33*(1), 42-49.

C-8 Educational Level by Age and Sex for Population Age 7 and Above Table For India- (Census-2011). Available from: .www.censusindia.gov.in/2011census/C-series/c-8/DDW-0000C-8, downloaded on 11/ 3 /2017.

Garson GD (2012) Testing statistical assumptions, pp: 44-45.

Global Adult Tobacco Survey (GATS) 2009-10, Available from URL http:// http://www.searo.who.int/tobacco/documents/2010-pub2.pdf, downloaded on 11/ 1 /2017.

Gupta, P. C., Maulik, P. K., Pednekar, M. S., & Saxena, S. (2005). Concurrent alcohol and tobacco use among a middle-aged and elderly population in Mumbai. *National Medical Journal of India*, *18*(2), 88.

International Institute for Population Sciences (IIPS) and Macro International. 2007. National Family Health Survey (NFHS-3), 2005–06: India: Volume I & Volume II. Mumbai: IIPS.

Jackson, K. M., Sher, K. J., Wood, P. K., & Bucholz, K. K. (2003). Alcohol and tobacco use disorders in a general population: short-term and long-term associations from the St. Louis Epidemiological Catchment Area Study. *Drug and alcohol dependence*, *71*(3), 239-253.

Larsen, K., & Merlo, J. (2005). Appropriate assessment of neighborhood effects on individual health: integrating random and fixed effects in multilevel logistic regression. *American journal of epidemiology*, *161*(1), 81-88.

Mathur, C., Stigler, M. H., Perry, C. L., Arora, M., & Reddy, K. S. (2008). Differences in prevalence of tobacco use among Indian urban youth: the role of socioeconomic status. *Nicotine & Tobacco Research*, *10*(1), 109-116.

Pandey, A., Choe, M. K., Luther, N. Y., Sahu, D., & Chand, J. (1998). Infant and Child Mortality in India: National Family Health Survey Subject Reports Number 11. *International Institute of Population Sciences: Mumbai*.

Pinheiro, J. C., & Chao, E. C. (2006). Efficient Laplacian and adaptive Gaussian quadrature algorithms for multilevel generalized linear mixed models. *Journal of Computational and Graphical Statistics*, *15*(1), 58-81.

Pinilla, J., Gonzalez, B., Barber, P., & Santana, Y. (2002). Smoking in young adolescents: an approach with multilevel discrete choice models. *Journal of Epidemiology & Community Health*, *56*(3), 227-232.

Rani, M., Bonu, S., Jha, P., Nguyen, S. N., &Jamjoum, L. (2003). Tobacco use in India: prevalence and predictors of smoking and chewing in a national cross sectional household survey. *Tobacco control*, *12*(4).

Raudenbush, S. W., & Bryk, A. S. (1992). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Sage.

Ravishankar, T. L., &Nagarajappa, R. (2009). Factors attributing to initiation of tobacco use in adolescent students of Moradabad,(UP) India. *Indian journal of dental research*, *20*(3), 346.

Sanagou, M., Wolfe, R., Forbes, A., & Reid, C. M. (2013). Hospital-level associations with 30-day patient mortality after cardiac surgery: a tutorial on the application and interpretation of marginal and multilevel logistic regression. *BMC medical research methodology*, *12*(1), 28.

Sorensen, G., Gupta, P. C., &Pednekar, M. S. (2005). Social disparities in tobacco use in Mumbai, India: the roles of occupation, education, and gender. *American journal of public health*, *95*(6), 1003-1008.

Sreeramareddy, C. T., Kishore, P. V., Paudel, J., & Menezes, R. G. (2008). Prevalence and correlates of tobacco use amongst junior collegiates in twin cities of western Nepal: a cross-sectional, questionnaire-based survey. *BMC Public Health*, *8*(1), 97.

Subramanian, S. V., Nandy, S., Kelly, M., Gordon, D., & Smith, G. D. (2004). Patterns and distribution of tobacco consumption in India: cross sectional multilevel evidence from the 1998-9 national family health survey. *Bmj*, *328*(7443), 801-806.

Yawson, A. E., Baddoo, A., Hagan-Seneadza, N. A., Calys-Tagoe, B., Hewlett, S., Dako-Gyeke, P., & Kowal, P. (2013). Tobacco use in older adults in Ghana: sociodemographic characteristics, health risks and subjective wellbeing. *BMC public health*, *13*(1), 979.

WHO Report on the Global Tobacco Epidemic, 2011, Warning about the dangers of tobacco, Available from URL http://www.who.int/tobacco/global_report/2011/en/, downloaded on 24/4/2013.