2018

Modeling and Predicting Prevalence of Tobacco Use in Bangladesh

Begum, Munjila

University of Rajshahi

MODELING AND PREDICTING PREVALENCE OF TOBACCO USE IN BANGLADESH

A Dissertation
Submitted to the
University of Rajshahi in Partial
Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

By

Munjila Begum

DEPARTMENT OF STATISTICS
UNIVERSITY OF RAJSHAHI
RAJSHAHI

JUNE, 2018

MODELING AND PREDICTING PREVALENCE OF TOBACCO USE IN BANGLADESH

Ph.D

Researcher

MUNJILA BEGUM
Session: 2013-2014
Department of Statistics
University of Rajshahi,
Rajshahi

Superviser

Dr. Mst. Papia Sultana
Professor
Department of Statistics
University of Rajshahi,
Rajshahi

Department of Statistics
University of Rajshahi,
Rajshahi

June, 2018

Dr. Mst. Papia Sultana
Professor
Department of Statistics
University of Rajshahi,
Rajshahi-6205
Bangladesh

ড. মোছাঃ পাপিয়া সুলতানা
প্রফেসর
পরিসংখ্যান বিভাগ
রাজশাহী বিশ্ববিদ্যালয়, রাজশাহী- ৬২০৫
Mobile: 01914254227
Email: papia.stat@ymail.com

CERTIFICATE

I have the pleasure to certify that the thesis entitled "MODELING AND PREDICTING PREVALENCE OF TOBACCO USE IN BANGLADESH" is the original work of Munjila Begum. As far as I know, this is the candidate's own achievement and is not a conjoint work. She has completed this thesis under my direct guidance and supervision. I also certify that I have gone through the draft and final version of the thesis and found it satisfactory for submission to the Department of Statistics, University of Rajshahi in fulfillment of the requirements for the degree of Doctor of Philosophy in Statistics.

Rajshahi
June 28, 2018

Supervisor

(Dr. Mst. Papia Sultana)
Professor
Department of Statistics
University of Rajshahi,
Rajshahi, Bangladesh.

DECLARATION

I do hereby declare that the thesis entitled, "MODELING AND PREDICTING PREVALENCE OF TOBACCO USE IN BANGLADESH", submitted to the Department of statistics, University of Rajshahi for the of Doctor of Philosophy in Statistics is exclusively my own and original work. This work is carried out by me under the supervision and guidance of Professor Dr. Mst. Papia Sultana, Department of Statistics, University of Rajshahi, Rajshahi, Bangladesh. No part of it any form, has been submitted to any other University or Institution for any degree, diploma or other similar purposes.

Rajshahi
28/06/2018

Munjila Begum
Ph.D Research Fellow
Department of Statistics,
University of Rajshahi
Rajshahi, Bangladesh.

ABSTRACT

Tobacco consumption is one of the major preventable causes of death and disability globally. Bangladesh is one of the largest tobacco consuming countries in the world. The use of tobacco is high among male population in Bangladesh. Also the use of tobacco is increasing. Therefore this study aims to explore the prevalence, patterns and determinants of tobacco use among adults.

This study used data from Global Adult Tobacco Survey (GATS), 2010 that covered 9,629 Bangladeshi aged 15 years or above. The survey was based on a three-stage stratified cluster sample of household. Information of a total of 9629 adults has been analyzed in the present study.

Pattern and prevalence of tobacco consumption in Bangladesh are primarily presented in the form of frequency distributions and in prevalence rates. This study uses binary logistic regression model and multilevel logistic regression model for analyzing data to quantify the objectives of the study.

Prevalence of adults tobacco user was 44.05% in Bangladesh with 65.11% in male and 34.89% in female. Common significant predictors include sex, age, educational level and wealth Index. Male and older had a higher tendency to use tobacco products. Current tobacco smoking was significantly higher among male (p-value<0.001 and OR=44.17) than female. Females were more likely to use smokeless tobacco than males (odds ratio, OR=1.72). Adults with no education were more likely to use tobacco products in Bangladesh compared to others with tertiary education. Adults with the poorest wealth status were more likely to consume tobacco products in Bangladesh compared to those from richest wealth index. It applies a multilevel (two level) logistic regression analysis to draw valid conclusions about the effects of the selected determinants on tobacco consumption using GATS-2010 data which is a multistage stratified cluster data. Instead of single level logistic model, multilevel logistic regression model has been utilized since the data follow a hierarchical structure. Also the comparison between single and multilevel model has been done to investigate the necessity of multilevel effects. The findings suggest that sex, age, level of education and wealth index have significant multilevel effects on tobacco consumption. Manufactured cigarettes and betel quid with zarda are the most usable tobacco products in Bangladesh. Intervention to reduce or stop tobacco using should be directed towards the poor, older and people with lower education and women should be targeted for prevention of the use of tobacco. All the significant variables should be considered for developing suitable policies to reduce the consequences of tobacco use in Bangladesh.

ACKNOWLEDGEMENTS

All praise goes to almighty Allah who has given me the opportunity to do this research. It is a great pleasure for me to express my sincere appreciation to the Department of Statistics, University of Rajshahi for giving me the opportunity to do this research.

I want to mention the contribution of all those who have inspired, influenced and guided me to complete this research work successfully. First of all, I would like to express my heartiest gratitude to my venerable supervisor Professor, Dr. Mst. Papia Sultana. Throughout my study, she kindly read my works and provided me with insightful comments substantively and editorial advice. I am grateful to him for hes kind and sincere guidance and personal supervision all through my work.

I am also grateful to Professor, Dr. Md. Ayub Ali, Chairman, Department of Statistics, University of Rajshahi, for giving me the permission to work on this thesis and providing me all sources of facilities of this department.

I wish to acknowledge my gratitude to all my respected teachers of Statistics Department for their suggestions and kind cooperation. Specially, I am very grateful to Dr. Md. Nurul Islam, Professor, Department of Statistics, University of Rajshahi.

Many of my friends deserve thanks for their encouragement and cooperation. My special thanks go to my youngest brothers Rakib, Rashed and Rana. I want to specially thank Dr. Aziza Sultana Rozy Sarkar, Lecturer, Department of statistics, Rajshahi College, Rajshahi and Eva, M.Sc. thesis student.

I would like to thank the members of the Evaluation Committee for their valuable comments and suggestions in several occasions namely, proposal and candidature defense and final presentation at the faculty. Finally, and most

importantly, I would like to thank to my family. None of this would have been possible without their love and support. I owe my deepest gratitude to my parents, sister, husband, Md. Ramjan Ali and son, Muntakim. Their understanding and love were undeniably the most important support for me.

Munjila Begum
PhD Research Fellow
Department of Statistics,
University of Rajshahi,
Rajshahi, Bangladesh.

June, 2018

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

INTRODUCTON

1.1 Background of the study

1.2 Literature review

1.3 Objectives of the study

1.4 Organization of the study

CHAPTER ONE

INTRODUCTION

1.1 Background of the study

Tobacco use is one of the leading preventable causes of premature death, disease and disability around the world. An estimated 4.9 million deaths occurring annually can be attributed to tobacco use. This figure is expected to rise to about 10 million by the year 2020 [http://www.who.int/whr/2002/en/], if the current trend continues unchecked, there will be up to one billion tobacco-related deaths during the 21st century, many of which will be from developing countries. [http://www.who.int/tobacco/mpower/tobacco_facts/en/index.html.]. Tobacco and poverty together form a vicious circle from which it is often difficult to escape. The adverse effects of tobacco use, including loss of income, being a leading causes of death, and contributing to chronic disease, are well documented worldwide (Ezzati M et al., 2003).The prevalence of tobacco use is an important predictor of the future burden of tobacco-related diseases (Flora MS et al., 2009). Despite scientific evidence linking tobacco to adverse health outcomes, a certain proportion of youth and adults are very likely to continue smoking, chewing and snuffing tobacco products unless there is continued public health action to control their use. Misperceptions about the dangers of using tobacco may partly explain why youth and adults use tobacco in developing countries (Elton-Marshall T et al., 2010). In Bangladesh, tobacco use has become not only a major contributor to the country's high morbidity but also the biggest drains to the nation's economy (Zamman MM et al., 2007). Bangladesh, like many transitional nations in the world is straddling through the demographic and epidemiological transitions. The country is observing large decline in mortality due to acute, infectious, and parasitic diseases and increases in chronic diseases such as heart disease and diabetes (collectively known as non-communicable diseases or NCDs) over the last 20 years. Consistent with this shifting epidemiological profile, recent data suggests that tobacco consumption is one of the most important modifiable risk factors contribute to these emerging chronic diseases (WHO, 2002) and has also been identified as a major risk factor for mortality. Bangladesh is ranked among the top tobacco consuming countries in the world (http://bdnews24.com/health/2014/12/17/ Bangladesh -among-top-smokeless-tobacco-using-nations). Form of smoked tobacco include manufactured cigarettes, bidis, pipes, cigars, water pipes and smokeless tobacco is usually consumed orally or nasally, without burning or combustion. The use increases the risk of cancer and leads to nicotine addiction similar to that produced by cigarette smoking or other smoked tobacco products. Different 12 types of smokeless tobacco are: chewing tobacco (most prevalent in Indian subcontinent); snuff (most prevalent in Scandinavian and US but becoming popular worldwide); and dissolvable (most prevalent in high income countries). The chewing tobacco is an oral smokeless tobacco product that is placed in the mouth, cheek, or inner lip and sucked or chewed. They are also referred to as “spit tobacco” because of the tendency by users to spit out the built-up tobacco juices and saliva. Overall, the percentage of Bangladeshi people who use smoked tobacco, smokeless tobacco or both, increased from 36.8% in 2004-05 to 43.2% in 2009 (ITC Bangladesh Summary, Promoting Evidence-Based Strategies to Fight the Global Tobacco Epidemic, 2010). This alarming rise poses severe impact on the country’s overall disease burden. A WHO study estimated that in 2004, 57 000 people lost their life prematurely as a result of tobacco use and 382,000 people became disabled in Bangladesh (D. Efroymson et al., 2001). In Bangladesh, the numbers of tobacco smokers are increasing rapidly because of the availability of cheap tobacco products, lack of strong tobacco control regulations, and weak enforcement of existing regulations. The Global Adult Tobacco Survey conducted by WHO reported that Bangladesh is one of the top ten countries in the world with high tobacco use (both smoking and smokeless forms) with a prevalence of 43.3% among adults (41.3 million) (http://www.who.int/tobacco/surveillance/survey/gats/en/). Smoking tobacco is a risk factor for several diseases and has been increasing in many developing countries. It is not only a global public health concern, but also an economic problem amongst individuals, societies, and the country as a whole. The risks of cancer, cardiovascular disease, respiratory disease, and a range of other health problems are increased in tobacco smokers and, as a consequence, smokers are more likely than nonsmokers to die prematurely. Smoking is considered a leading cause of morbidity and mortality in virtually every country in the world, and it is the second only to high blood pressure as a risk factor for global disease burden. Tobacco use causes more than 440,000 deaths in the US per year, accounting for one out of every five deaths. In addition, up to two-thirds of deaths in current smokers can be attributed to smoking. The higher prevalence of tobacco use in the developing countries are anticipated to result in large disease burden in the near future. Tobacco and poverty together form a vicious circle from which it is often difficult to escape. The adverse effects of tobacco use, including loss of income, being a leading causes of death, and contributing to chronic disease, are well documented worldwide. The prevalence of tobacco use is an important predictor of the future burden of tobacco-related diseases. It is estimated that each year tobacco smoking accounts for about 9% of deaths globally. Around 71% of lung cancer, 42% of chronic respiratory diseases, and nearly 10% of cardiovascular diseases are caused by smoking. It is reported that 18% of deaths in high-income countries have occurred due to tobacco use, whereas in middle- and low-income countries it is 11% and 4% respectively (Giovino GA et al., 2012). In low- and middle-income countries such deaths are projected to increase from 3.4 to 6.8 million between 2002 and 2030 (Hu T-w et al., 2005). In addition, secondhand smoke exposure poses a serious risk of causing heart disease and various respiratory illness, lung cancer, etc among nonsmokers. Both smoking and chewing tobacco products are commonly used in Bangladesh. The smokeless tobacco use constitutes a major part of overall tobacco use in Bangladesh and India (Palipudi KM et al., 2015). Smoking tobacco products include cigarettes, bidis (a small, thin, hand-rolled cigarette consisting of tobacco leaf, manufactured mostly in India and Bangladesh), hookah (a water pipe which is used to smoke tobacco through cooled water). Chewing or smokeless tobacco products include betel quid with tobacco (also known as pan, which is a mixture of betel leaf, areca nut, slaked lime, and tobacco), zarda (a mixture of tobacco, lime, spices, and vegetable dyes), zarda with areca nut, and gul (an oral tobacco powder that is rubbed over the gum and teeth). It is found that 28.30% men and 0.20% women in Bangladesh smoke cigarettes. In the Indian subcontinent, poor people use bidis as smoking tobacco. It has also been documented that the main predictors of cigarette smoking are sex, age, and having friends who smoke (Reda AA et al., 2012). Moreover, cigarette smoking is considered as a “gate way” toward illegal drug use, especially among adolescents. Several strategies have been shown to reduce tobacco use. However, more than 50 years after the health dangers of smoking were scientifically proven, and more than 20 years after evidence confirmed the hazards of second-hand smoke, few countries have implemented effective and recognized strategies to control the tobacco epidemic. International efforts led by WHO resulted in rapid entry into force of the WHO Framework Convention on Tobacco Control (WHO FCTC), which has 168 signatories and more than 150 Parties. Achievement of tobacco control goals will require coordination among many government agencies, academic institutions, professional associations and civil society organizations at the country level, as well as the coordinated support of international cooperation and development agencies. Various socioeconomic factors are found to be associated with different types of tobacco use. Studies regarding tobacco use in developing countries provided mixed results (Bush J et al., 2003). Thus, it is an important task to identify the determining factors of tobacco use in Bangladesh.

1.2 Literature Review

Some research works already have done in related area. Some of the previous empirical studies as well as of other countries in this area has been conducted. Citable research works are summarized below:

Achia TNO (2015) examined the association between self-reported tobacco use and frequency of mass media utilization by women and men in nine low-to middle-income sub-Saharan African countries. Data for the study came from Demographic and Health

Surveys conducted in Burkina Faso, Ethiopia, Liberia, Lesotho, Malawi, Swaziland, Uganda, Zambia and Zimbabwe over the period 2006–2011. Each survey population was a cross-sectional sample of women aged 15–49 years and men aged 15–59 years, with information on tobacco use and media access being obtained by face-to-face interviews. An index of media utilization was constructed based on responses to questions on the frequency of reading newspapers, frequency of watching television and frequency of listening to the radio. Demographic and socioeconomic variables were considered as potentially confounding covariates. Logistic regression models with country and cluster specific random effects were estimated for the pooled data. The risk of cigarette smoking increased with greater utilization to mass media. The use of smokeless tobacco and tobacco use in general declined with greater utilization to mass media. The risk of tobacco use was 5% lower in women with high media utilization compared to those with low media utilization [Adjusted Odds Ratio (AOR) = 0.95, 95% confidence interval (CI):0.82–1.00]. Men with a high media utilization were 21% less likely to use tobacco compared to those with low media utilization [AOR = 0.79, 95%CI = 0.73–0.85]. In the male sample, tobacco use also declined with the increased frequency of reading newspapers (or magazines), listening to radio and watching television.

Aziz S, Choudhury T and Huque N (2015) explored the association of tobacco consumption with socio-demographic factors, self-rated health and non-communicable chronic diseases among the rural population in Bangladesh. A cross sectional data from “IUB Health and Socio Economic Survey” 2013 was used for this study. A sample of 1512 male and 1569 female aged 18 and above were randomly selected from four districts of Bangladesh. Binary logistic regression was used to explore the association of tobacco consumption with socio-demographic characteristics; self reported chronic diseases and general health. The prevalence of current tobacco use in any form (smoking or chewing tobacco) among adults in rural Bangladesh was 47.4%. Poor self reported general health was found to be a significant predictor of tobacco consumption. Males were 2.13 times more likely to use tobacco than their female counterparts. Tobacco use was significantly associated with older age (OR=3.18, CI=

2.507-4.035), higher education (OR=0.426, CI=0.289-0.628) and Sylhet region (OR=3.707, CI=2.911-4.720).

Bhise MD and Patra S (2018) estimated the current prevalence of hypertension and its correlates in the state of Maharashtra. The variation in the prevalence of hypertension associated with individual-level characteristics is explained at the community and district level. Data was used from the recent round of District Level Household & Facility Survey (DLHS-4), 2012–13. The DLHS-4 had used the nationally representative sample, collected through multistage stratified sampling procedure. A similar sampling frame, used in NSSO-2007-08, had been followed. The chi-square test was used to show the significance level of the association between the estimated prevalence of hypertension and its correlates. Multilevel regression analysis was carried out to investigate the effects of individual and community level factors on the prevalence of hypertension. The overall prevalence of hypertension is 25% in Maharashtra, and a huge variation in the prevalence of hypertension is found across the districts. Dhule, Gadchiroli (with a low HDI rank), Mumbai and Satara (with higher HDI rank) were the districts with the higher (above 30%) prevalence of high blood pressure. The prevalence also significantly varied according to different correlates. The prevalence of high blood pressure is higher among elderly population (40%), among males (28%), in the urban areas (27%) and in the richest wealth quintile (28%). The prevalence was also higher among cigarette smokers (31%), alcohol consumers (30%) and people with obesity (38%) as compared to their counterparts. The results of the multilevel analysis showed that the older and obese persons were at four-time higher risk of hypertension. Again, age, sex, marital status, place of residence, wealth status, unhealthy habits (i.e. smoking and alcohol consumption) and BMI were significantly associated with hypertension. The results of VPC statistics show that 14% of hypertension prevalence could be attributed to differences at the community level.

Corsi DJ, Lear SA, Chow CK, Subramanian SV, Boyle MH and Teo kk (2013) described the socioeconomic and geographic distribution of smoking behaviour in Canada among 19,383 individuals (51% women) aged 15–85 years. Current smoking and quitting were modeled using standard and multilevel logistic regression. Markers of socioeconomic status (SES) were education and occupation. Geography was defined by Canadian Provinces. Results: The adjusted prevalence of current smoking was 20.2% (95% confidence interval [CI]: 18.8–21.7) and 63.7% (95% CI: 61.1–66.3) of ever smokers had quit. Current smoking decreased and quitting increased with increasing SES. The adjusted prevalence of current smoking was 32.8% (95% CI: 28.4–37.5) among the least educated compared to 11.0% (95% CI: 8.9– 13.4) for the highest educated. Among the least educated, 53.0% (95% CI: 46.8–59.2) had quit, rising to 68.7% (95% CI: 62.7– 74.1) for the most educated. There was substantial variation in current smoking and quitting at the provincial level; current smoking varied from 17.9% in British Columbia to 26.1% in Nova Scotia, and quitting varied from 57.4% in Nova Scotia to 67.8% in Prince Edward Island. Nationally, increasing education and occupation level were inversely associated with current smoking (odds ratio [OR] 0.64, 95% CI: 0.60–0.68 for education; OR 0.82, 95% CI: 0.77–0.87 for occupation) and positively associated with quitting (OR 1.27, 95% CI: 1.16–1.40 for education; OR 1.20, 95% CI: 1.12–1.27 for occupation). Those associations were consistent in direction across provinces although with some variability in magnitude.

Dhungana RR, Khanal MK and Baniya A (2013) conducted meta-analysis to estimate the prevalence of current tobacco use among lower secondary to higher secondary students in Nepal. The study searched and identified the studies which were published between 2003 and 2013 using MEDLINE, Google Scholar and NEPJOL. From five selected studies, total 7,832 eligible students were included in analysis. Considering the high degree of variability (Q = 82.6, I2 = 95%) among selected studies, the study used random effects model to estimate the weighted prevalence of current tobacco use and found as 13.9 % (10.2-17.5). This result showed that current tobacco use among lower secondary to higher secondary students still remains high, which compels an effective implementation of tobacco control programs and policies.

Do YK and Bautista MA (2015) aimed to investigate the associations between tobacco use within households and expenditures on food, education, and healthcare in LMICs. Using data from the World Health Survey , this cross-sectional study included a sample of 53,625 adult males aged <60 years from 40 LMICs. Multilevel, mixed-effects linear regression was used to determine the association between current tobacco use status of the main income provider (daily; occasional; no use) and three categories of (logged) household expenditures: food, education, and healthcare; controlling for age, level of education, household wealth quintile, marital status, urban–rural setting, country-level income group, and region. Results In the preferred random-slope models that controlled for covariates, daily tobacco use was associated with lower household expenditures on education and healthcare by 8.0 % (95 % confidence interval:(−12.8 to –3.2 %) and 5.5 % (−10.7 to –0.3 %), respectively . The association between tobacco use and food expenditure was inconsistent across models. Conclusions T obacco use in LMICs may have a negative influence on investment in human capital development. Addressing the tobacco use problem in LMICs could benefit not only the health and economic well-being of smokers and their immediate families but also long- run economic development at a societal level.

Evans-Whipp TJ, Bond L, Ukoumunne OC, Toumbourou JW and Catalano RF (2010) measured tobacco polices in statewide representative samples of secondary and mixed schools in Victoria, Australia and Washington, US (N = 3,466 students from 285 schools) and tested their association with student smoking. Results from confounder-adjusted random effects (multi-level) regression models revealed that the odds of student perception of peer smoking on school grounds were decreased in schools that have strict enforcement of policy (odds ratio (OR) = 0.45; 95% CI: 0.25 to 0.82; p = 0.009). There was no clear evidence in this study that a comprehensive smoking ban, harsh penalties, remedial penalties, harm minimization policy or abstinence policy impact on any of the smoking outcomes.

Evans-Whipp TJ, Bond L, Ukoumunne OC, Toumbourou JW and Catalano RF (2010) measured tobacco polices in statewide representative samples of secondary and mixed schools in Victoria, Australia and Washington, US (N = 3,466 students from 285 schools) and tests their association with student smoking. Results from confounder-adjusted random effects (multi-level) regression models revealed that the odds of student perception of peer smoking on school grounds were decreased in schools that had strict enforcement of policy (odds ratio (OR) = 0.45; 95% CI: 0.25 to 0.82; p = 0.009). There was no clear evidence in this study that a comprehensive smoking ban, harsh penalties, remedial penalties, harm minimization policy or abstinence policy impact on any of the smoking outcomes.

Gaete J, Ortúzar C, Zitko P, Montgomery A and Araya R (2016) examined the inter-school variability in student smoking in a large sample of Chilean schools and identifies the school- and student-level characteristics associated with cigarette smoking This cross-sectional study used self-reported student-level data from 45,273 students from 1462 schools and official data from these schools provided by the Chilean Ministry of Education (2007). Student smoking behavior was used as an outcome, and individual-level and school-level features were used as explanatory variables. Logistic multilevel modeling was used to analyses the data. The mean prevalence of smoking in the 1462 schools was 39.9 %. The null model indicated that 8 % of the variance in smoking behavior was explained by schools; and in the final model, controlled by individual- and school-level variables, the variance explained by schools dropped to 2.4 %. The main school-level variables explaining the school influence were school bonding, school truancy and school achievement.

Gilani SI and Leon DA (2013) conducted a cross-sectional survey of nationally representative sample of men and women living in rural and urban areas of four main provinces of Pakistan from March through April 2012. Face to face in house interviews were undertaken using a pre-tested structured questionnaire that asked about smoking and other forms of tobacco use. Multistage stratified random area probability sampling was used. To determine the national prevalence of tobacco use, the sample was weighted to correspond to rural-urban population proportions in each of the four provinces as in the 1998 census conducted by Pakistan's Population Census Organization. Association between socio-demographic variables and tobacco use were investigated using multivariable robust regression. Out of 2,644 respondents (1354 men and 1290 women), 345 men and 4 women reported being current cigarette smokers. The weighted prevalence of current cigarette smoking was 15.2% (95% CI:

11.2, 19.3) overall, 26.6% (95% CI: 19.1, 34.1) among male, and 0.4% (95% CI: -0.2,1.0) among females. Among females 1.8% (95% CI: 0.4, 3.1) used any smoked tobacco and 4.6% (95% CI: 1.8,7.4) used any smokeless tobacco daily or on some days of the week. Among males, odds of current cigarette smoking decreased with increasing level of education (OR=0.76; 95% CI: 0.68, 0.84) and increased with having a father who used tobacco (OR=2.11; 95% CI: 1.39,3.22) after adjusting for other socio-demographic characteristics. Lower household income was associated with current cigarette smoking among rural males only (odds ratio [OR] =0.67; 95% CI: 0.48, 0.92 per category increase in monthly household income).

Guo Q, Johnson CA, Unger JB, Lee L, Xie B, Chou CP, Palmer PH, Sun P, Gallaher P and Pentz MA (2007) explored whether the Theory of Reasoned Action (TRA) and Theory of Planned Behavior (TPB) predict adolescent smoking in China. Data were obtained from 14,434 middle and high school students (48.6% boys, 51.4% girls) in seven geographically varied cities in China. TRA and TPB were tested by multilevel mediation modeling, and compared by multilevel analyses and likelihood ratio tests. Perceived behavioral control was tested as a main effect in TPB and a moderation effect in TRA. The mediation effects of smoking intention were supported in both models (pb0.001). TPB accounted for significantly more variance than TRA (pb0.001). Perceived behavioral control significantly interacted with attitudes and social norms in TRA (pb0.001). Therefore, TRA and TPB were applicable to China to predict adolescent smoking. TPB was superior to TRA for the prediction and TRA could better predict smoking among students with lower than higher perceived behavioral control.

Hossain MS, Kypri K, Rahman B, Arsian I, Akter S and Milton AH (2014) estimated the prevalence and identify correlates of smokeless tobacco consumption among married rural women with a history of at least one pregnancy in Madaripur, Bangladesh. The study conducted a cross-sectional survey using an interviewer administered, pre-tested, semi-structured questionnaire. All women living in the study area, aged 18 years and above with at least one pregnancy in their lifetime, who were on the electoral roll and agreed to participate were included in the study. Information on socio-demographic characteristics and smokeless tobacco consumption was collected. Smokeless tobacco consumption was categorized as 'current', 'Ever but not current' and 'Never'. Associations between smokeless tobacco consumption and the explanatory variables were estimated using simple and multiple binary logistic regression.

Huang H-L, Chen F-L, Hsu C-C, Yen Y-Y, Ted Chen, Huang C-M, Shi H-Y, Hu C-Y and Lee C-H (2010) examined school based tobacco policy status, implementation and students' perceived smoking at school in regard to gender-specific differences in smoking behavior. The study conducted a multilevel-based study to assess two-level effects for smoking among 2350 grades three to six students in 26 randomly selected elementary schools in southern Taiwan. A series of multilevel models were analyzed separately for male and female students. The school level variables appear to be related to smoking behavior in male students. Among males, the risk of ever-smoking was significantly associated with those schools without antitobacco health education activities or curricula [adjusted odds ratio (aOR: 56.23, 95% confidence interval (CI): 2.55–15.24), with a high perceived smoking rate (aOR53.08, 95% CI: 1.41–6.72) and located in a mountainous region (aOR: 52.53, 95% CI: 1.15–5.58). The risk of ever-smoking among females was significantly associated with those schools without antitobacco activities or curricula (aOR 53.10, 95% CI: 1.27–7.55). As compared with female counterparts, the specific school that the male students attended had a positive significant effect on the risk of being ever-smokers. The findings suggest that effective tobacco policy implementation should be considered in elementary schools that were currently putting children at the greatest risk for cigarette smoking, especially in regard to male students.

Imtiaz D, Kandpal SD and Juyal R (2015) assessed the quitting behavior among current tobacco users in a rural population of Dehradun. The study was cross sectional in nature carried out among 993 current tobacco users aged 10 years and above in the field practice area and quitting behavior was assessed using a pretested and predesigned questionnaire. Of the 993 Current tobacco users, 38% and 40% of the current smokers and current smokeless tobacco users respectively had attempted to quit smoking and smokeless tobacco use in the past 12 months. 54.3% of the smokers wanted to quit smoking with majority of male smokers (56.3%) willing to quit smoking compared to only 40.5% of female smokers. 36.0% of the smokeless tobacco users wanted to quit smokeless tobacco use where in contrast more female smokeless tobacco users (39.3%) wanted to quit smokeless tobacco compared to 33.3% of males.

Jafarabadi MA, Allahverdipour H, Bashirian S, and Jannati A (2012) studied underlying factors in predicting the behavior of tobacco smoking among employed youth and students in Iran. In this analytical cross-sectional study, based a random cluster sampling were recruited 850 high school students, employed and unemployed youth age ranged between 14 and 19 yr from Iran. The data of demographic and tobacco smoking related variables were acquired via a self-administered questionnaire. A series of univariate and multivariate logistic regressions were performed respectively for computing un-adjusted and adjusted Odds Ratios utilizing SPSS 17 software.

Khan MH, Khan A, Kraemer A and Mori M (2009) used secondary data which was collected by the 2006 Urban Health Survey. The data were representative for the urban areas in Bangladesh. Both slums and non-slums located in the six City Corporations were considered. Slums in the cities were identified by two steps, first by using the satellite images and secondly by ground truthing. At the next stage, several clusters of households were selected by using proportional sampling. Then from each of the selected clusters, about 25 households were randomly selected. Information of a total of 12,155 adult men, aged 15–59 years, was analyzed by stratifying them into slum (= 6,488) and non-slum (= 5,667) groups. Simple frequency, bivariable and multivariable logistic regression analyses were performed using SPSS. Overall smoking prevalence for the total sample was 53.6% with significantly higher prevalence among men in slums (59.8%) than non-slums (46.4%). Respondents living in slums reported a significantly ($P < 0.001$) higher prevalence of smoking cigarettes (53.3%) as compared to those living in non-slums (44.6%). A similar pattern was found for bidis (slums = 11.4% and non slums = 3.2%, $P < 0.001$). Multivariable logistic regression revealed significantly higher odds ratio (OR) of smoking cigarettes (OR = 1.12, 95% CI = 1.03–1.22), bidis (OR = 1.90, 95% CI = 1.58–2.29) and any of the two (OR = 1.23, 95% CI = 1.13–1.34) among men living in slums as compared to those living in non-slums when controlled for age, division, education, marital status, religion, birth place and types of work. Division, education and types of work were the common significant correlates for both cigarette and bidi smoking in slums and non-slums by multivariable logistic regressions. Other significant correlates of smoking cigarettes were marital status (both areas), birth place (slums), and religion (non-slums). Similarly significant factors for smoking bidis were age (both areas), marital status (slums), religion(non-slums), and birth place (both areas).

Khan MHR and Shaw JE (2011) examines the selected determinants of contraceptive prevalence among 10-49 aged ever-married women in Bangladesh and their true impact on the contraception prevalence rate (CPR). It applied a multilevel logistic regression analysis to draw valid conclusions about the effects of the selected determinants on CPR using the 2004 Bangladesh Demographic and Health Survey (BDHS) contraceptive binary data which wass a multistage stratified cluster data. Instead of standard single level logistic model, multilevel logistic regression model had been utilized since the data followed a hierarchical structure. Also the comparison between single and multilevel model had been done to investigate the necessity of multilevel effects. The findings suggested that age of the women, number of living children, education, religion, media, place of residence and wealth index had significant multilevel effects on CPR. The study had finally suggested integrating a strong awareness program that targets the 10-49 aged currently married women in Bangladesh in those divisions (level-3) and clusters (level-2) where the particular determinant of contraceptive use had been found to be less effective.

Khayyati F, Taymoori P, Mohammadpoorasl A, Allahverdipour H and Asghari Jafarabadi M (2016) studied underlying predictors of tobacco smoking among Iranian Teenagers in a generalized structural equation model. In this cross-sectional study, a Generalized Structural Equation Model based on planned behavioral theory was used to explain the relationship among different factors such as demographic factors, subjective norms, and the intention to tobacco and, in turn, intention with tobacco use. The sample consisted of 4,422 high school students, based on census, in East Azerbaijan province, Iran. The questioner was designed adapting to the objectives of study. It was used global youth tobacco survey to design the queries of tobacco use. The model had a good fit on data. Adjusting for age and gender, there was a statistically significant relationship between the intention to consumption and the following factors: working while studying ($P<0.05$), consumption of hookah by family members ($P<0.05$), history of smoking among close friends ($P<0.05$), history of leaving school during the day without informing the officials ($P<0.05$), lack of commitment to academic tasks ($P< 0.05$), lack of acceptance by peers ($P< 0.05$), knowledge ($P< 0.05$), and attitude ($P< 0.05$). There was a significant relationship between the intentions to consumption to tobacco consumption in the past 30 days ($P< 0.05$) as well.

Kishore J, Jena PK, Bandyopadhyay C, Swain M, Das S and Banerjee I (2013) assessed the prevalence and associated factors of hardcore smoking in three South-East Asian countries and discussed its implication for smoking cessation intervention in this region. Global Adult Tobacco Survey (GATS) data of India, Bangladesh and Thailand were analyzed to quantify the hardcore smoking prevalence in the region. On the basis of review, an operational definition of hardcore smoking was adopted that includes (1) current daily smoker, (2) no quit attempt in the past 12 months or not interested in quitting, (4) time to first smoke within 30 minutes of waking up, and (5) knowledge of smoking hazards. Logistic regression analysis was carried out using hardcore smoking status as response variable and gender, type of residence, occupation, education, wealth index and age-group as possible predictors. There were 31.3 million hardcore smokers in the three Asian countries. The adult prevalence of hardcore smoking in these countries ranges between 3.1% in India to 6% in Thailand. These hardcore smokers constitute 18.3-29.7% of daily smokers. The logistic regression model indicated that age, gender, occupation and wealth index were the major predictors of hardcore smoking with varied influence across countries.

Kristjansson AL, Sigfusdottir ID, and Allegrante JP (2013) sought to add to a growing body of literature into peer contexts by testing a model of peer substance use simultaneously on individual and school community levels while taking account of several well established individual level factors. We analyzed population based data from the 2009 Youth in Iceland school survey, with 7,084 participants (response rate of 83.5%) nested within 140 schools across Iceland. Multilevel logistic regression models were used to analyze the data. School level peer smoking and drunkenness were positively related to adolescent daily smoking and lifetime drunkenness after taking account of individual level peer smoking and drunkenness. These relationships held true for all respondents, irrespective of socio-economic status and other background variables, time spent with parents, academic performance, self-assessed peer respect for smoking and alcohol use, or if they had substance-using friends or not. On the other hand, the same relationships were not found with regard to individual and peer cannabis use. The school-level findings in the study represent context effects that were over and above individual-level associations. This holds although accounted for a large number of individual level variables that studies generally had not included. For the purpose of prevention, school communities should be targeted as a whole in substance use prevention programs in addition to reaching to individuals of particular concern.

Leatherdale ST and Manske S (2005) examined how perceptions of student smoking in the school environment and the actual smoking rate among senior students at a school are related to smoking onset. Multilevel logistic regression analysis was used to examine correlates of ever smoking in a sample of 4,286 grade 6 and 7 students from 57 elementary schools in Ontario, Canada. Students are at increased risk for smoking if they (a) often see students smoking near their school, (b) report that students at their school smoke where they are not allowed, and (c) attend a school with a relatively high senior student smoking rate. Each 1% increase in the smoking rate among grade 8 students increased the odds that a student in grades 6 or 7 was an ever smoker versus never smoker (odds ratio, 1.05; 95% confidence interval, 1.02-1.08). A low-risk student (no family or friends who smoke) was over twice as likely to try smoking if he/she attended a high-risk school. Prevention programs should target both at-risk schools and at-risk students, and strongly enforced policies preventing students from smoking on or near school property should be implemented.

Manimunda SP, Benegal V, Sugunan AP, Jeemon P, Balakrishna N, Thennarusu K, Pandian D and Pesala KS (2012) Conducted a cross-sectional survey among a representative sample of 18,018 individuals in the age group of >=14 years in the Union Territory of Andaman and Nicobar Islands during 2007-09. A structured questionnaire, a modified version of an instrument which was used successfully in several multi-country epidemiological studies of the World Health Organization, was used to survey individual socio-demographic details, known co-morbid conditions, tobacco use and alcohol use. Fagerstrom Test for Nicotine Dependence (FTND) was used to estimate nicotine dependence. The response rate of our survey was 97% (18,018/18,554). Females (n=8,888) were significantly younger than males (34.3 ± 14.6 Vs 36.2 ± 15.4 years). The prevalence of current tobacco use in any form was 48.9% (95% CI: 48.2-49.6). Tobacco chewing alone was prevalent in 40.9% (95% CI: 40.1-41.6) of the population. While one ten of males (9.7%, 95% CI: 9.1-10.4) were nicotine dependent, it was only 3% (95% CI: 2.7-3.4) in females. Three fourth of the tobacco users initiated use of tobacco before reaching 21 years of age. Age, current use of alcohol, poor, educational status, marital status, social groups, and co-morbidities were the main determinants of tobacco use and nicotine dependence in the population.

Merlo J, Wagner P, Ghith N and Leckie G (2016) proposed an original stepwise analytical approach that distinguishes between “specific” (measures of association) and “general” (measures of variance) contextual effects. Performing two empirical examples they illustrated the methodology, interpreted the results and discussed the implications of this kind of analysis in public health. They analysed 43,291 individuals residing in 218 neighbourhoods in the city of Malmö, Sweden in 2006. We study two individual outcomes (psychotropic drug use and choice of private vs. public general practitioner, GP) for which the relative importance of neighbourhood as a source of individual variation differs substantially. In Step 1 of the analysis, they evaluated the OR and the area under the receiver operating characteristic (AUC) curve for individual-level covariates (i.e., age, sex and individual low income). In Step 2, they assessed general contextual effects using the AUC. Finally, in Step 3 the OR for a specific neighbourhood characteristic (i.e., neighbourhood income) was interpreted jointly with the proportional change in variance (i.e., PCV) and the proportion of ORs in the opposite direction (POOR) statistics. For both outcomes, information on individual characteristics (Step 1) provide a low discriminatory accuracy (AUC = 0.616 for psychotropic drugs; = 0.600 for choosing a private GP). Accounting for neighbourhood of residence (Step 2) only improved the AUC for choosing a private GP (+0.295 units). High neighbourhood income (Step 3) was strongly associated to choosing a private GP (OR = 3.50) but the PCV was only 11% and the POOR 33%.

Moore L, Roberts C, Tudor-Smith C (2001) examined the association between school smoking policies and smoking prevalence among pupils. Multilevel analysis of crosssectional data from surveys of schools and pupils, 55 secondary schools in Wales, 55 teachers and 1375 pupils in year 11 (aged 15–16) were computed. Main outcome measures was self-reported smoking behavior. The prevalence of daily smoking in schools with a written policy on smoking for pupils, teachers, and other adults, with no pupils or teachers allowed to smoke anywhere on the school premises, was 9.5% (95% confidence interval (CI) 6.1% to 12.9%). In schools with no policy on pupils' or teachers' smoking, 30.1% (95% CI 23.6% to 36.6%) of pupils reported daily smoking. In schools with an intermediate level of smoking policy, 21.0% (95% CI 17.8% to 24.2%) smoked every day. School smoking policy was associated with school level variation in daily smoking (p = 0.002). In multilevel analysis, after adjusting for pupils' sex, parents' and best friends' smoking status, parental expectations, and alienation from school, there was less unexplained school level variation, but school smoking policy remained significant (p = 0.041). The association of smoking policy with weekly smoking was weaker than for daily smoking, and not significant after adjustment for pupil level variables. Both daily and weekly smoking prevalence were lower in schools where pupils' smoking restrictions were always enforced. Enforcement of teacher smoking restrictions was not significantly associated with pupils' smoking.

Palipudi KM, Gupta PC, Sinha DN, Andes LJ, Asma S and McAfee T (2012) examined the role of social determinants on current tobacco use n thirteen low-and-middle income countries. The study used nationally representative data from the Global Adult Tobacco Survey (GATS) conducted during 2008-2010 in 13 low and middle income countries: Bangladesh, China, Egypt, India, Mexico, Philippines, Poland, Russian Federation, Thailand, Turkey, Ukraine, Uruguay, and Viet Nam. These surveys provided information on 209,027 respondent's aged 15 years and above and the country datasets were analyzed individually for estimating current tobacco use across various socio-demographic factors (gender, age, place of residence, education, wealth index, and knowledge on harmful effects of smoking). Multiple logistic regression analysis was used to predict the impact of these determinants on current tobacco use status. Current tobacco use was defined as current smoking or use of smokeless tobacco, either daily or occasionally. Former smokers were excluded from the analysis. Adjusted odds ratios for current tobacco use after controlling other cofactors, was significantly higher for males across all countries and for urban areas in eight of the 13 countries. For educational level, the trend was significant in Bangladesh, Egypt, India, Philippines, Thailand, Turkey, Ukraine, Uruguay and Viet Nam. The trend of decreasing prevalence with increasing levels of knowledge on harmful effects of smoking was significant in China, India, Philippines, Poland, Russian Federation, Thailand, Ukraine and Viet Nam.

Palipudi KM, Sinha DN, Choudhury S, Zaman MM, Asma S, Andes L and Dube S (2015) examined predictors of current tobacco smoking and smokeless tobacco use among the adult population in Bangladesh. The study used data from the 2009 Global Adult Tobacco Survey (GATS) in Bangladesh consisting of 9,629 adults aged≥15 years. Differences in and predictors of prevalence for both smoking and smokeless tobacco use were analyzed using selected socioeconomic and demographic characteristics that included gender, age, place of residence, education, occupation, and an index of wealth. The prevalence of smoking was high among males (44.7%, 95% confidence interval [CI]: 42.5‑47.0) as compared to females (1.5%, 95% CI: 1.1‑2.1), whereas the prevalence of smokeless tobacco was almost similar among both males (26.4%, 95% CI: 24.2‑28.6) and females (27.9%, 95% CI: 25.9‑30.0). Correlates of current smoking were male gender (odds ratio [OR] =41.46, CI=23.8‑73.4), and adults in older age (ORs range from 1.99 in 24‑35 years age to 5.49 in 55‑64 years age), less education (ORs range from 1.47 in less than secondary to 3.25 in no formal education), and lower socioeconomic status (ORs range from 1.56 in high wealth index to 2.48 in lowest wealth index. Predictors of smokeless tobacco use were older age (ORs range from 2.54in 24‑35 years age to 12.31 in 55‑64 years age), less education (ORs range from 1.44 in less than secondary to 2.70 in no formal education), and the low (OR=1.34, CI=1.0‑1.7) or lowest (OR=1.43, CI=1.1‑1.9) socioeconomic status.

Park S, Nam B-H, Yang H-R, Lee JA, Lim H, Han JT, Park Su, Shin HR and Lee JS (2013) developed an individualized risk prediction model for lung cancer in Korean men using population-based cohort data. From a population-based cohort study of 1,324,804 Korean men free of cancer at baseline, the individualized absolute risk of developing lung cancer was estimated using the Cox proportional hazards model. We checked the validity of the model using C statistics and the Hosmer–Lemeshow chi-square test on an external validation dataset. The risk prediction model for lung cancer in Korean men included smoking exposure, age at smoking initiation, body mass index, physical activity, and fasting glucose levels. The model showed excellent performance (C statistic = 0.871, 95% CI = 0.867–0.876). Smoking was significantly associated with the risk of lung cancer in Korean men, with a four-fold increased risk in current smokers consuming more than one pack a day relative to non-smokers. Age at smoking initiation was also a significant predictor for developing lung cancer; a younger age at initiation was associated with a higher risk of developing lung cancer.

Pinilla J, González B, Barber P and Santana Y (2002) estimated the effects of individual, family, social, and school related factors. Cross sectional analysis performed by multilevel logistic regression with pupils at the first level and schools at the second level. The data came from a stratified sample of students surveyed on their own, their families’ and their friends’ smoking habits, their schools, and their awareness of cigarette prices and advertising. The study was performed in the Island of Gran Canaria, Spain. Participants:1877 students from 30 secondary schools in spring of 2000 (model's effective sample sizes 1697 and 1738) . 14.2% of the young teenagers surveyed use tobacco, almost half of them (6.3% of the total surveyed) on a daily basis. According to the ordered logistic regression model, to have a smoker as the best friend increases significantly the probability of smoking (odds ratio: 6.96, 95% confidence intervals (CI) (4.93 to 9.84), and the same stands for one smoker living at home compared with a smoking free home (odds ratio: 2.03, 95% CI 1.22 to 3.36). Girls smoke more (odds ratio: 1.85, 95% CI 1.33 to 2.59). Experience with alcohol, and lack of interest in studies are also significant factors affecting smoking. Multilevel models of logistic regression showed that factors related to the school affect the smoking behavior of young teenagers. More specifically, whether a school complies with antismoking rules or not was the main factor to predict smoking prevalence in schools. The remainder of the differences could be attributed to individual and family characteristics, tobacco consumption by parents or other close relatives, and peer group. A great deal of the individual differences in smoking were explained by factors at the school level. The most relevant predictors for smoking in young adolescents included some factors related to the schools they attend. One variable stood out in accounting for the school to school differences: how well they enforced the no smoking rule. Therefore we could prevent or delay tobacco smoking in adolescents not only by publicizing health risks, but also by better enforcing no smoking rules in schools.

Piontek D, Buehler A, Donath C, Floeter S, Metz URK, Gradl S and Kroeger C (2008) applied a multilevel approach to examine the associations between school smoking policy and student smoking. It was tested whether individual characteristics are mediators of school policy effects. On the basis of cross-sectional data from 3,364 students and school principals from 40 schools in Germany, two multilevel nonlinear regression models were computed for current smoking. In the first model, controlling for individual factors not influenced by school, smoking bans for students and evidence-based prevention activities were negatively associated with smoking prevalence. The second model included student characteristics potentially influenced by school (e.g. school engagement, peer smoking). As school variables remained significant, these characteristics obviously do not mediate school context effects.

Rahman MM, Karim MJ, Ahmad SA, Suhaili MR and Ahmad SNW (2011) determined the prevalence of smoking and to examine the determinants of smoking behavior among the secondary school teachers in Bangladesh. A two-stage cluster sampling was used with a selection of schools on Probability Proportional to Enrolment (PPE) size followed by stratified random sampling of government and private schools and then all the teachers present on the day of the survey were selected for the study. The 66-item questionnaire included smoking behavior, knowledge, attitude, second-hand smoking, tobacco free school policy, cessation, media advertisement and curriculum related topics. Seven additional questions were included to assess the socio-demographic characteristics of the teachers. Data analysis was performed using SPSS 17 software. A total of 60 schools were selected with school response rate of 98.3%. An anonymous self-administered questionnaire was filled in by all teachers present at the day of the survey. The sample consisted of 559 teachers with response rate of 99.5%. The prevalence of smoking was 17% (95% CI: 14%, 20.4). About half of the teachers (48.4%) smoke daily followed by 25.3% smoke 1-2 days in last 30 days. The mean duration smoking of was 13.7 (95% CI: 11.6, 15.9) years. Logistic regression analysis revealed that male teachers smoke 37.46 (95% CI: 5.078, 276.432) times higher than their female counterparts. The graduate teachers were 2.179(95% 1.209, 3.926) times more likely to be smoke than master's degree holder teachers. Smoking by friends appeared to be the strongest predictor for teachers smoking behavior (OR 4.789, 95% CI: 1.757, 13.050). However, no statistically significant association was found between type of school, second-hand smoking and curriculum related factors and smoking behavior of the teachers ($p>0.05$).

Rahman MS, Mondal MNI, Islam MR, Rahman MM, Hoque MN and Alam MS (2015) explored the different types of tobacco use, and to identify the determinant factors associated with the tobacco use among ever-married men in Bangladesh. Data of 3,771 ever-married men, 15–54 years of age were extracted from the Bangladesh

Demographic and Health Survey 2007. Prevalence rate, chi-square (X^2) test, and binary logistic regression analysis were used as the statistical tools to analyze the data. Tobacco use through smoking (58.68%) was found to be higher than that of chewing (21.63%) among men, which was significantly more prevalent among the poorest, less educated, and businessmen. In bivariate analysis, all the socioeconomic factors were found significantly associated with tobacco use; while in multivariate analysis, age, education, wealth index, and occupation were identified as the significant predictors. Tobacco use was found to be remarkably common among males in Bangladesh.

Rani M, Bonu S, Jha P, Nguyen SN and Jamioum L (2003) estimated the prevalence and the socioeconomic and demographic correlates of tobacco consumption in India. Cross sectional, nationally representative population based household survey was used. 315598 individuals 15 years or older from 91196 households were sampled National Family Health Survey-2 (1998-99). Data on tobacco consumption were elicited from household informants. Prevalence of current smoking and current chewing of tobacco were used as outcome measures. Simple and two way cross tabulations and multivariate logistic regression analysis were the main analytical methods. Thirty percent of the population 15 years or older-47% men and 14% of women-either smoked or chewed tobacco, which translates to almost 195 million people-154 million men and 41 million women in India. However the prevalence may be underestimated by almost 11% and 1.5% for chewing tobacco among men and women, respectively and by 5% and 0.5% for smoking among men and women, respectively, because of use of household informants. Tobacco consumption was significantly higher in poor, less educated scheduled castes and scheduled tribe populations. The prevalence of tobacco consumption increased up to the age of 50 years and then leveled or declined. The prevalence of smoking and chewing also varied widely between different states and had a strong association with individual's sociocultural characteristics.

Siahpush M, Mbiostat, Spittal M and Singh GK (2007) examined the association of smoking cessation with financial stress and material well-being. Data (n = 5699 at baseline) came from 4 consecutive waves (2001–2005) of the Household Income and

Labour Dynamics in Australia survey. The study used mixed models to examine the participant-specific association of smoking cessation with financial stress and material well-being. On average, a smoker who quits was expected to have a 25% reduction ($P<.001$; odds ratio [OR] = 0.75; 95% confidence interaval [CI] = 0.69, 0.81) in the odds of financial stress. Similarly, the data provided strong evidence ($P<.001$) that a smoker who quits is likely to experience an enhanced level of material well-being. The findings indicated that interventions to encourage smoking cessation are likely to improve standards of living and reduce deprivation. The findings provided grounds for encouraging the social services sector to incorporate smoking cessation efforts into their programs to enhance the material or financial conditions of disadvantaged groups.

Sinha DN, Palipudi KM, Rolle I, Asma S and Rinchen S (2011) examined the prevalence of current tobacco use among youth and adults in selected member countries of the South-East Asia Region using the data from school and household-based surveys included in the Global Tobacco Surveillance System. Global Youth Tobacco Survey (GYTS) data (years 2007-2009) were used to examine current tobacco use prevalence among youth, whereas Global Adult Tobacco Survey (GATS) data (years 2009-2010) were used to examine the prevalence among adults. GYTS is a school-based survey of students aged 13-15, using a two-stage cluster sample design, and GATS is a household survey of adults age 15 and above using a multi-stage stratified cluster design. Both surveys used a standard protocol for the questionnaire, data collection and analysis. Prevalence of current tobacco use among students aged 13-15 varied from 5.9% in Bangladesh to 56.5% in Timor-Leste, and the prevalence among adults aged 15 and above was highest in Bangladesh (43.3%), followed by India (34.6%) and Thailand (27.2%). Reported prevalence was significantly higher among males than females for adults and youth in all countries except Bangladesh, Sri Lanka and Timor-Leste. Current use of tobacco other than manufactured cigarettes was notably higher than current cigarette smoking among youth aged 13-15 years in most countries of the Region, while the same was observed among adults in Bangladesh, India and Thailand, with most women in those countries, and 49% of men in India, using smokeless tobacco.

Spitz MR, Hong WK, Amos CI, WU X, Schabath MB, Dong Q, Shete S and Etzel CJ (2007) constructed and validated a comprehensive clinical tool for lung cancer risk prediction by smoking status. Epidemiologic data from 1851 lung cancer patients and 2001 matched control subjects were randomly divided into separate training (75% of the data) and validation (25% of the data) sets for never, former and current smokers, multivariable models were constructed from the training sets. The discriminatory ability of the models was assessed in the validation sets by examining the areas under the receiver operating characteristic curves and with concordance statistics. Absolute 1-year risks of lung cancer were computed using national incidence and mortality data. An ordinal risk index was constructed for each smoking status category by summing the odds ratios from the multivariate regression analyses for each rick factor. All variables that had a statistically significant association with lung cancer (environmental tobacco smoke, family history of cancer, dust exposure, prior respiratory disease and smoking history variables) had strong biologically plausible etiologic roles in the disease. The concordance statistics in the validation sets for the never, former and current smokers model were 0.57, 0.63 and 0.58, respectively. The computed 1-year absolute risk of lung cancer for a hypothetical male current smoker with an estimated relative risk close to 9 was 8.68%. The ordinal risk index performed well in that true positive rates in the designated high-risk categories were 69% and 70% of current and former smokers, respectively.

Sultana P, Akter S, Rahman MM and Alam MS (2015) estimatd the prevalence and identified the socioeconomic and demographic correlates of tobacco smoking in Bangladesh. Secondary data had been used of size 9629 (male=4468 and female=5161) aged 15 years and above collected by the Global Adult Tobacco Survey (GATS), 2010. Principle Component Analysis (PCA) had been used to get the wealth index. Binary logistic regression model has been used to assess the predictors of current tobacco smoking. Prevalence of current tobacco smokers in Bangladesh was 23.19 (48.28% male and 1.47% female) and the prevalence of current daily tobacco smokers was 21.16. Rural respondents were significantly more likely to smoke tobacco currently. Comparative to females, males were more likely to smoke currently (adjusted OR=37.55, 95% CI=25.91, 54.41). Respondents of youngest age group (15-24 yrs) were less likely to smoke currently than all other age groups and respondents with no formal schooling were more likely to smoke than respondents with all other levels of education. Businessmen, farmers and workers/day labours were more likely to smoke, and employee, students and respondents with other jobs were less likely to smoke. It was also found that respondents with lowest wealth index were most likely to smoke and respondents with higher wealth index were least likely to smoke. The results revealed that in Bangladesh, tobacco smoking was strongly associated with social disadvantage, for example, low socio-economic status, less education, stressed or low-paid job, etc.

1.3 Aims and objectives of the study

Tobacco is widely used and man made product. It is the primary cause of preventable illness and premature death. Though it offers us a life of slaver, a host of chronic, debilitating illness and ultimately death, the use of tobacco is not stopped. From literature review it is observed that few works already have been done in the related field (Rani M et al. 2003, Zorrilla-Torras B. 2005, Gupta PC et al. 2000, Pedenekar MS et al. 2008 and Sorensen G et al. 2005). But still tobacco related death is increasing and no serious anti-smoking efforts are made. The present study is motivated to fulfill this aspect.

The specific objectives of the research are:

- To investigate the pattern of tobacco use in Bangladesh.
- To examine the prevalence and predictors of tobacco consumption (Smoked and Smokeless) among adults in Bangladesh.

1.4 Organization of the study

The study has been carried out the following seven chapters maintaining the proper sequence.

Chapter One: Introduction with background, the review of earlier studies, aims and objectives have been discussed in this chapter.

Chapter Two: Chapter two presents the data source including study design. Methodology of the study has also been discussed in this chapter, too.

Chapter Three: In this chapter characteristics of the study subject has been represented and some of their corresponding graphs have been presented. Descriptive analysis has been performed to know the characteristics of the study subjects. For that frequencies with percentages have been reported.

Chapter Four: In chapter four pattern and prevalence of tobacco use (smoked, smokeless and both) among adults in Bangladesh have been reported. Current pattern of using tobacco products have been reported with prevalence and 95% confidence interval. P-values are obtained from Z-test for proportion. These test have been performed at 5% level of significance. Age adjusted and unadjusted prevalence have been reported also.

Chapter Five: In chapter five, Reporting Odds Ratio and 95% CI of predictors of tobacco using from Binary Logistic Regressions among Adults by Socio-demographic and economic characteristics.

Chapter Six: Chapter six provides the empirical findings on tobacco consumption among adults. The analysis covers the suitable techniques for using tobacco in different settings. Multilevel Logistic regression analysis have been performed on current tobacco consumption (smoked, smoking and smokeless) daily and compared with the result obtained from logistic regression analysis. Measures of association (odds ratio) and measures of variance (intra-class correlation (ICC), Median Odds Ratio (MOR)) were calculated, as well as the discriminatory accuracy by calculating the area under the ROC curve (AUC). Interpretation of parameters have been done in terms of level of significance (p-value=0.05) and odds ratio.

Chapter Seven: Chapter seven summarizes various findings of the study and concludes about the findings, contributions and limitations of the study and stating the policy implications necessary to decline the use of tobacco in Bangladesh.

CHAPTER TWO

DATA AND METHODOLOGY

2.1 Introduction

2.2 Data source

2.3 Sampling design

2.4 Questionnaire

2.5 Methodology

2.6 Data screening

CHAPTER TWO

DATA AND METHODOLOGY

2.1 Introduction

In chapter one the background and objectives of the study have been discussed. It also reviews some important and relevant literature. In any study it is essential to mention the data source, to discuss and narrate the methodology of the study. In this chapter the data source, the study design and a general description of Questionnaire (information on tobacco use) have been discussed.

2.2 Data Source

For our study, we have used the data from Global Adult tobacco Survey (i.e. GATS: Version 2.0. Atlanta, GA: Centers for Disease Control and Prevention, 2010). The survey was conducted in Bangladesh, Brazil, China, Egypt, India, Mexico, Philippines, Poland, Russia, Thailand, Turkey, Ukraine, Uruguay and Vietnam from 2008 to 2010. We will use only the data of Bangladesh which have information on 9629 respondents aged 15 years and above. The Global Adult Tobacco Survey (GATS) is a nationally representative household survey of men and women aged 15 years and above. It is designed to produce internationally comparable data on tobacco use and tobacco control measures using a standardized questionnaire, sample design, data collection, aggregation and analysis procedures. In Bangladesh, the survey was implemented by the National Institute of Preventive and Social Medicine (NIPSOM) with the collaboration of National Institute of Population Research and Training (NIPORT) and the Bangladesh Bureau of Statistics (BBS). The Centers for Disease Control and Prevention (CDC), United States, and the World Health

Organization provided technical assistance. [Global Adult Tobacco Survey, 2010]

2.3 Sampling design

The sampling frame used for GATS, Bangladesh was the population census of the People's Republic of Bangladesh conducted by Bangladesh Bureau of Statistic (BBS). The survey was based on a three-stage stratified cluster sample of households. At the first stage 400 Primary Sampling Units (PSU)s (Mauza in rural and Mohalla in urban areas) were selected with probability proportional to size (PPS), followed by a random selection of one Secondary Sampling Unit (SSU) per selected PSU. At the third stage households were selected systematically within the listed households from a selected SSU. SSUs were based upon Enumeration Areas (EA) from the Bangladesh Agricultural Census. These selected EAs were updated with mapping and listing. Typically these EAs consisted of 200 household units in Mauzas and 300 household units in each Mohalla. The explicit stratification used at the first stage of selection based upon urban (Mahalla) and rural (Mauza) designation of BBS. Each list of rural and urban geopolitical units was implicitly stratified by division, and within division by the percent literacy of women in each Mahalla and Mauza. Data were collected from 200 urban and 200 rural primary sampling units (mauza in rural and mohalla in urban areas). Sample design for Bangladesh consists of 400 PSUs, 200 in urban areas and 200 in rural areas. After accounting for possible non-response and eligibility rates, it was determined to have an average of 28 households per selected SSU resulting in a total sample size of 11200 non-institutionalized households from all 6 administrative divisions covering 95.5% of the total population. As per design, one respondent was randomly selected for the interview from each selected eligible household to participate in the survey. The Bangladesh sample design provides cross-sectional estimates for the country as a whole as well as by urban, rural and gender.

Details about the study design will be found in Global Adult Tobacco Survey Collaborative Group. 2010a. *Global Adult Tobacco Surveys (GATS): Sample Design Manual. [http://* www.cdc.gov/tobacco/global/gats.]

From the Figure 2.1 below we observed the hierarchical structure of the data.

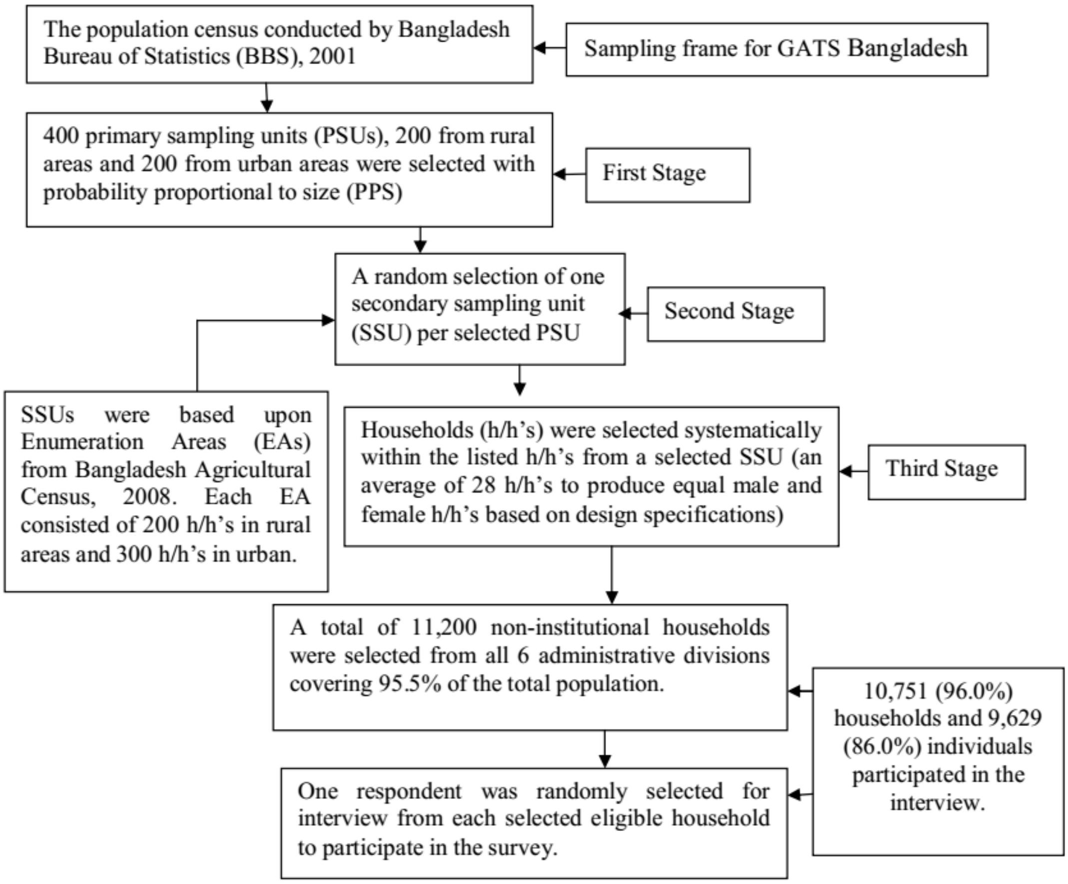

Figure 2.1: Sampling design of GATS in Bangladesh.

2.4 Questionnaire

A general description of the GATS Bangladesh questionnaire is described below.

- **Background characteristics:** Gender, Residence, age, education, work status, wealth index.
- **Tobacco smoking:** Patterns of use (daily consumption, less than daily consumption, not at all), former/past tobacco consumption, age of initiation of daily smoking, consumption of different tobacco products, (cigarettes, pipes, cigars and other smoked tobacco), nicotine dependence, frequency of quit attempts.
- **Smokeless tobacco:** Patterns of use (daily consumption, less than daily consumption, not at all), former/past use of smokeless tobacco, age of initiation of daily use of smokeless tobacco, consumption of different smokeless tobacco products (snuff, chewing tobacco, betel quid, etc.), nicotine dependence, frequency of quit attempts.
- **Cessation:** Advice to quit smoking by health care provider, method used to try to stop smoking. Similar questions were included about cessation of smokeless tobacco use as well.
- **Secondhand smoke (SHS):** Smoking allowed in the home; exposure to SHS at home; indoor smoking policy at the workplace; exposure in last 30 days in: workplace, government buildings/offices, health care facilities, restaurants, public transportation. There were some additional optional items on exposure that included schools, universities, private workplaces, bars, night clubs, etc., as well as knowledge about serious illness in non-smokers due to SHS.
- **Economics:** Type of tobacco product and quantity bought, cost of tobacco product (s), brand and type of product purchased, and source of tobacco products.
- **Media:** Exposure to tobacco advertising on television, radio, billboards, posters, newspapers/magazines, cinema, Internet, public transportation, public walls, and others; exposure to sporting events connected with tobacco; exposure to music, theatre, art or fashion events connected with tobacco; exposure to tobacco

promotion activities; reaction to health warning labels on cigarette 30 packages; and exposure to anti-tobacco advertising and information. Similar questions were included for smokeless tobacco as well. The reference period for the questions in this section was 30 days.

- **Knowledge, attitudes and perceptions:** Knowledge about the health effects of both smoking and smokeless tobacco.

Details about the questionnaires can be found in Global Adult Tobacco Survey Collaboration Group, 2010. Global Adult Tobacco Survey (GATS). Core Questionnaire with Optional Questions, version 2.0, Atlanta, GA: Centers for Diseases Control and Prevention. http://www.cdc.gov/tobacco/global/gats**.**

2.5 Methodology

Various statistical methodology have been used to analyze the data. Descriptive statistics have been computed to obtain the basic information about the respondents. For that frequencies with percentages have been reported.

Frequencies indicate the number of cases (respondents), which falls into each of the available categories. Frequencies can be displayed in terms of counts or percentages. Frequencies are usually displayed by means of frequency tables, but can also be displayed graphically in graphs and charts. Suitable graphs to display frequencies for categorical data are bar charts or pie charts.

Graphical representation of a frequency distribution is more effective than tabular representation and it is also easily comprehensible. Diagram is essential to convey the statistical information to the general public. It also facilitates the comparison of two or more frequency distribution. Data presented in the form of tables give good information in concise form. Tables provide all relevant information of the data. Apart from tabular presentation, graphical presentation of data has also become quite popular. It gives visual information in addition to magnitudes. Furthermore, comparisons and changes in the data can be well visualized when presented in graphical form. A very useful part of graphical presentation is the interpretation of the graphs. In every graph we should try to interpret the data. But these may broadly be categorized into the following:

– Bar chart
– Pie chart
– Histogram
– Frequency polygon
– Pareto chart
– Frequency curve
– Line diagram.

Bar charts are used for categorical data or metric data that are transformed into categorical data. Categories are shown on the horizontal axis. Frequency, percentage, or proportion is shown on the vertical axis. Bars are separated from each other to emphasize the distinctness of the categories. The bars must be of the same width. The length of each bar is proportional to the frequency, percentage, or proportion in the category. Levels ought to be provided on both axes.

Like bar charts, **pie charts** are also used for categorical data. A circle is divided into segments, the areas of which are proportional to the values in the question. But the areas are proportional to the angles the corresponding segments make at the center of the circle. Thus, segments of the circle are cut in such a way that their values are proportional to the angles. **Histograms** are used for metric data but converted to categories. These are somewhat similar to bar charts. However, there are some important features in histograms. The blocks in histograms are placed together one after another. These are not separated. Classes are ordered on the horizontal axis, with scores increasing from left to right. Areas of the blocks are proportional to the frequencies. If the class intervals are of equal width, the heights of the blocks/rectangles are proportional to the frequencies. If the class intervals are of unequal width, the blocks/rectangles are drawn in such a way that the areas of the blocks/rectangles are proportional to the frequencies. However, it is easier to interpret the histograms, if the class intervals are of equal width.

Frequency Polygon is also a graphical presentation of frequency distribution. It is more convenient than the histogram. The midpoints of the upper extremes of the blocks of the histogram are joined by straight lines. The first and the last parts of the polygon are to be brought to the horizontal axis at a distance equal to half of the class width. A **pareto chart** is a bar chart for count (discrete) data. It displays the frequency of each count on the vertical axis and the count type on the horizontal axis. The count types are always arranged in descending order of frequency of occurrence. The most frequent occurring type is on the left, followed by the next–most frequently occurring type, and so on. Bars are placed side by side with no gap between the adjacent ones. A segmented line is also drawn to depict the relative cumulative frequency distribution. Pareto charts are useful, among other uses, in the analysis of defect data in manufacturing system, construction management, and others, and is an important part of quality improvement program since it allows the management and engineers to focus attention on the most critical defects in a production or process. Line diagrams are drawn by plotting the values of two continuous variables. These show trends or changes in one variable resulting from changes in the other. One important application of the line diagram is to study the changes of various economic indicators over time. Line diagrams may be presented in the form of continuous lines or segmented lines depending on the phenomenon under study. Frequency curve is a smoothed frequency polygon. It is produced by plotting the absolute frequency of an infinitesimally small range of a continuous variable. It is a theoretical distribution.

A test of proportion will assess whether or not a sample from a population represents the true proportion from the entire population. The steps to perform a test of proportion using the critical value approval are as follows:

(1) State the null hypothesis H_0 the alternative hypothesis H_A.

(2) Calculate the test statistic:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

Where p_0 is the null hypothesized proportion i.e., when H_0: $p=p_0$

(3) Determine the critical region.

(4) Make a decision. Determine if the test statistic falls in the critical region. If it does, reject the null hypothesis. If it does not, do not reject the null ypothesis.

To test for significance of association between variables, the Chi-square tests were used. The null hypothesis of no relationship between the dependent variable and the determinant variable is rejected if the p-value of the Chi-square ($\chi 2$) statistic is less than 5%. The relationships between variables were estimated using regressions. Logistic regression is useful when the response variable is categorical in nature. For comparison, statistical testing procedure such as Chi-square test, and compute p-value. P-value tells us whether the variables are statistically and significantly associated or not. If the p-value is smaller than 0.05, we know the results are statistically significant at 5% level.

Chi Square Test

The "theoretical frequency" for a cell, given the hypothesis of independence, is

$$E_{i,j} = \frac{\left(\sum_{n_c=1}^{c} O_{i,n_c}\right)\cdot\left(\sum_{n_r=1}^{r} O_{n_r,j}\right)}{N}$$

where N is the total sample size (the sum of all cells in the table).

The value of the test-statistic is

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(O_{i,j} - E_{i,j}\right)^2}{E_{i,j}}$$

where,

χ^2 =Pearson's cumulative test statistic, which asymptotically approaches a χ^2 distribution.

$O_{i,j}$ = an observed frequency;

$E_{i,j}$ = an expected (theoretical) frequency, asserted by the null hypothesis;

n = the number of cells in the table.

Fitting the model of "independence" reduces the number of degrees of freedom by p = r + c − 1. The number of degrees of freedom is equal to the number of cells rc, minus the reduction in degrees of freedom, p, which reduces to (r − 1)(c − 1). Where r is the number of levels for one categorical variable, and c is the number of levels for the other categorical variable.

The approximation to the chi-squared distribution breaks down if expected frequencies are too low. It will normally be acceptable so long as no more than 20% of the events have expected frequencies below 5. Where there is only 1 degree of freedom, the approximation is not reliable if expected frequencies are below 10. In this case, a better approximation can be obtained by reducing the absolute value of each difference between observed and expected frequencies by 0.5 before squaring; this is called Yates's correction for continuity. In cases where the expected value, E, is found to be small (indicating either a small underlying population probability, or a small number of observations), the normal approximation of the multinomial distribution can fail, and in such cases it is found to be more appropriate to use the G-test, a likelihood ratio-based test statistic. Where the total sample size is small, it is necessary to use an appropriate exact test, typically either the binomial test or (for contingency tables) Fisher's exact test; but note that this test assumes fixed and known marginal totals.

In statistics, **G-tests** are **likelihood-ratio or maximum likelihood** statistical significance tests that are increasingly being used in situations where chi-squared tests were previously recommended.

The general formula for G is

$$G = 2\sum_i O_i \cdot \ln\left(O_i / E_i\right),$$

where O_i is the observed frequency in a cell, E_i is the expected frequency on the null hypothesis, and the sum is taken over all cells, and where $\ln$ denotes the natural logarithm(log to the base *e*) and the sum is taken over all non-empty cells.

Fisher's exact test is a statistical significance test used in the analysis of contingency tables. Although in practice it is employed when sample sizes are small, it is valid for all sample sizes. It is named after its inventor, R. A. Fisher, and is one of a class of exact tests, so called because the significance of the deviation from a null hypothesis can be calculated exactly, rather than relying on an approximation that becomes exact in the limit as the sample size grows to infinity, as with many statistical tests.

In statistics, **Yates's correction for continuity** (or **Yates's chi-squared test**) is used in certain situations when testing for independence in a contingency table. In some cases, Yates's correction may adjust too far, and so its current use is limited. Using the chi-squared distribution to interpret Pearson's chi-squared statistic requires one to assume that the discrete probability of observed binomial frequencies in the table can be approximated by the continuous chi-squared distribution. This assumption is not quite correct, and introduces some error. To reduce the error in approximation, Frank Yates, an English statistician, suggested a correction for continuity that adjusts the formula for Pearson's chi-squared test by subtracting 0.5 from the difference between each observed value and its expected value in a 2×2 contingency table. This reduces the chi-squared value obtained and thus increases its p-value. The effect of Yates's correction is to prevent overestimation of statistical significance for small data. This formula is chiefly used when at least one cell of the table has an expected count smaller than 5. Unfortunately, Yates's correction may tend to overcorrect. This can result in an overly conservative result that fails to reject the null hypothesis when it should (a type II error).

So it is suggested that Yates's correction is unnecessary even with quite low sample sizes, such as:

$$\sum_{i=1}^{N} O_i = 20$$

The following is Yates's corrected version of Pearson's chi-squared statistic:

$$\chi^2_{Yates} = \sum_{i=1}^{N} \frac{\left(\left|O_i - E_i\right| - 0.5\right)^2}{E_i}$$

where

O_i = an observed frequency
E_i = an expected (theoretical) frequency, asserted by the null hypothesis
N= number of distinct events.

Logistic Regression

Many tobacco related studies have employed logistic regression in their analysis (Abdullah et al., 2011; Chen et al., 2009; Khan et al., 2009; Palipudi et al., 2012; Rahman et al., 2011). The main reason is to deal with variables that are categorical. Linear regression cannot deal with dependent variables that are categorical in nature and the alternatives are a number of regression techniques, including logistic regression (Agresti, 2007). Frequently "logistic regression" refers to the technique for problems in which the dependent variable is dichotomous (the category of dependent variable is limited to two categories). When there are more than two categories, the techniques are referred to as multinomial logistic regression and if the multiple categories are ordered, then ordinal logistic regression is used (Bender & Grouven, 1997; Chan, 2004). Logistic regression was used to predict the outcome of a categorical dependent variable based on one or more independent variables. The categories or groups of the dependent variables are mutually exclusive and exhaustive, that is a case can only be in one group or the other, and every case must be a member of one of the groups. Large sample sizes are recommended since maximum likelihood estimators for the coefficients are large sample approximations. Therefore, the recommendation for logistic regressions should be at least 50 cases per predictor (Bender & Grouven, 1997).The maximum likelihood (or ML) estimation is used to fit the model. To test the significance of the logistic regression, two hypotheses are of interest (a) null hypothesis, which states that all the coefficients in the regression equation take the value zero, and (b) alternative hypothesis that the model with predictors currently under consideration is significant and differs from the null value of zero, i.e. is considerably better than the chance or random prediction. The log likelihood (LL) is the basis for tests of a logistic model and is based on –2LL ratio. This is a test of significance of difference between the likelihood ratio (–2LL) for the investigator’s model with predictors minus the likelihood ratio for the baseline model with only a constant in it (Chan, 2004). The Hosmer and Lemeshow (H-L) test for binary logistic regression, and Pearson and Deviance tests for multinomial and ordinal logistic regressions are used to measure the goodness of fit of the model. Moreover, the Pseudo R2 (Cox & Snell, Nagelkerke, and McFadden) statistics are for measuring the strength of association between the dependent and independent variables. The overall classification accuracy (in percent) shows the percentage of cases that are correctly classified by the model (Chan, 2004). However, in ordinal logistic regression, tests of parallel lines are used with the assumption that the relationships between the predictor variables and the logits are same for all the logits (Bender & Grouven, 1997) Logistic regression is useful for situations in which we want to be able to predict the presence or absence of a characteristic or outcome based on values of a set of predictor variables. It is similar to a linear regression model but is suited to models where the dependent variable is dichotomous. Logistic regression coefficients can be used to estimate odds ratios for each of the independent variables in the model. Logistic regression is applicable to a broader range of research situations than discriminant analysis. Logistic regression is more applicable because of its distribution free assumption of the categorical independent variable. Interpretation of logistic regression can be done in terms of odd ratio, which may be the fundamental reason why logistic regression has proven such a powerful tool for epidemiologic research. In the analysis of dichotomous outcome or response variable many distribution functions have been proposed for use. Discussing some of these, Cox suggested the logistic distribution. The logistic distribution preferred for two primary reasons:

1. From a mathematical point of view, it is an extremely flexible and easily used function.
2. It lands itself to a biologically meaningful interpretation.

Logistic regression model consider a categorical variable (dichotomous variable) as of dependent variable.

Let Y is a dichotomous dependent variable, which take value 0 and 1. Thus

$$Y_i = \begin{cases} 1, & \text{if the individual uses tobacco during the study} \\ 0, & \text{otherwise} \end{cases}$$

where, i = 1,2,..........................n.

Also consider a collection of k independent variables which will be denoted by the vector

$X' = (X_1, X_2, \ldots\ldots, X_k)$ and β be a $(K+1)\times 1$ vector of unknown parameters. For simplification, we will use the quantity $\pi(X) = P(Y=1|X)$ the probability that the event occurs conditional on the value of X.

$$P(Y=1|X) = \pi(X_i) = \frac{e^{g(x_i)}}{1+e^{g(x_i)}} = \frac{e^{x_i g}}{1+e^{x_i g}} \qquad \ldots\ldots\ldots\ldots(2.1)$$

And

$$P(Y=0|X) = 1-\pi(X_i) = \frac{1}{1+\exp(x_i\beta)} \qquad \ldots\ldots\ldots\ldots(2.2)$$

The central part of logistic regression in a transformation of $\pi(X)$ is known as logit transformation. Which is defined in terms of $\pi(X)$, as follows:

$$g(X_i) = \log it\,\pi(X_i) = \log\left[\frac{\pi(x_i)}{1-\pi(x_i)}\right] = x_i\beta$$

$$\Rightarrow g(x_i) = \beta_0 + \beta_1 x_{i1} + \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots + \beta_k x_{ik} \qquad \ldots\ldots\ldots\ldots(2.3)$$

which is the logit of the multiple logistic regression models. The logit, $g(X)$ is linear in its parameters and has many of the desirable properties of linear regression model. The logit, $g(X)$ may be continuous and depending on the range of X it may range from -∞ to ∞ (Hosmer and Lemeshow, 1989).

Model Diagnostics

After estimating the Logistic regression model parameters using the maximum likelihood estimator, there is a need to assess the significance of the variables with regards to predicting the response variable. There are a number of statistics that can be used to carry out the assessment and these include deviance, likelihood ratio, Wald test and Score Test. These tests are discussed in the sections below.

For testing the significance of the parameters of logistic regression model following test procedures are usually used

1. Likelihood ratio test
2. Score test
3. Wald test

Likelihood ratio test is a general test procedure introduced by Neyman and Pearson in 1928 is known as the likelihood ratio test. This test is based on **maximum likelihood estimates**. The likelihood ratio test can be used for testing a sample or composite hypothesis against a simple or composite hypothesis.

In logistic regression, the likelihood ratio test used for testing the overall significance of coefficient for all the parameters. Our hypotheses are as follows

$$H_0 : \beta_1 = \beta_2 = \ldots\ldots\ldots\ldots \beta_p = 0$$

against, H_1 : At least one of them is not equal to zero

The likelihood ratio test is based on the ratio of two likelihood functions.

The comparison of observed values to predicted values using the likelihood function is based on the following expression:

$$D = -\sum \ln\left(\frac{L_0}{L_1}\right)$$

where, L_0 =Likelihood function for the current model.

L_1 =Likelihood function for the standard model

A standard model is one that contains as many parameters as there are in the data set. The quantity inside the brackets in the above expression is called the likelihood ratios and a test based on it is called likelihood ratio test. The statistic D is equation (β) is called the deviance by McCullagh and Nelder (1985) and plays a central rate in some approaches to assessment of goodness of fit. For assessing the significance of an independent variable we compare the value of D with and without the independent variable in the equation. The effect of including the independent variable in the model can be obtained by G as follows

G = D(for the model without the variable)-D(for the model with the variable)

i.e. G measure the change in D due to inclusion of the independent variable in the model, G can be expressed as

$$G = -\sum \log \left[\frac{(Liklihood \quad without \quad the \quad \text{var}iable)}{Liklihood \quad with \quad the \quad \text{var}iable)} \right]$$

Under the null hypothesis that $\beta_i's$ (i=1,2,........,p) are equal to zero, the statistics G follows chi-square distribution with p degrees of freedom. If the null hypothesis is rejected we may conclude that all the coefficient is not equal to zero i.e. at least on of the coefficient (β_i) has significant effect.

Robert F. Engle showed that the **Wald test**, the likelihood-ratio test and the Lagrange multiplier test (also known as the score test) are asymptotically equivalent.

We have used Wald test, Wald theory for testing the significance of the parameters of logistic regression model are discussed.

The Wald test procedure was introduced by Wald in 1943 and named according to his name.

A Wald test can be used in a great variety of different models including models for dichotomous variables and models for continuous variables. In logistic analysis due to the nature of maximum likelihood estimation Wald test has a definite advantage over the likelihood ratio test. But it has the same assumption as those of likelihood ratio test, when the overall null hypothesis $H_0 : \beta_1 = \beta_2 = \dots\dots\dots = \beta_p = 0$ is rejected then to identify the significant coefficient Wald test is used.

The Wald test is obtained by comparing the maximum likelihood estimate of any parameter to the estimate of its standard error. For testing

$$H_0 : \beta_i = 0$$

against, $H_1 : \beta_i \neq 0$ for $i = 0,1,2,\dots\dots, p$

the univariate Wald statistic is defined as

$$W_i = \frac{\hat{\beta}_i}{S \cdot E\left(\hat{\beta}_i\right)} \qquad \dots\dots\dots\dots\dots (2.4)$$

where $\hat{\beta}_i$ is maximum likelihood estimate of β_i and $S.E\left(\hat{\beta}_i\right)$ denotes the standard error of $\hat{\beta}_i$. Under the null hypothesis W_i follows a standard normal distribution.

The multivariate of the Wald test can be expressed as

$$W = \hat{\beta}'\left[\operatorname{var}\left(\hat{\beta}\right)\right]^{-1}\hat{\beta} \qquad \dots\dots\dots\dots\dots (2.5)$$

where, $\hat{\beta}$ is the maximum likelihood estimate of vector of parameter β and $\operatorname{var}\left(\hat{\beta}\right)$ is the estimated variance-covariance matrix, which is the inverse of the information matrix.

Under the null hypothesis

$$H_0 : \beta_1 = \beta_2 = \dots\dots\dots = \beta_p = 0$$

W follows Chi-square distribution with k degrees of freedom.

Wald test has the limitation that it behaved in an aberrant manner and often failing to reject the null hypothesis when the coefficient is significant, which was examined by Hauck and Donner (2012).

For testing whether the overall effect in significant or not the **score test** is used. We assumed that the asymptotic distribution of score vector is known and Cox and Hinkly (1974) showed that under the following regularity condition score vector $U(\beta)$ is asymptotically normally distributed with mean 0 and variance-covariance matrix $I(\beta)$.

The regularity conditions are;

1. The order of integration and differentiation are interchangeable.
2. The dimension of the parameter space Ω is finite and the value of the parameter is interior to Ω.
3. The probability distributions for different values of β are distinct.
4. The first two derivatives of log-likelihood with respect to β exist in the neighborhood of the true parameter value.

Under above regularity conditions the central limit theorem can be applied to the above score vector $U(\beta)$. As a result $U(\beta)$ follows asymptotically normally distributed with mean vector 0 and variance-covariance matrix $[I(\beta)]$.

For β be the (p+1) vector of parameters the hypothesis can be written as

$$H_0 : \beta = \beta_0$$

Against, $H_1 : \beta \neq \beta_0$

Under the null hypothesis the score statistic $U(\beta_0)$ is asymptotically normally distributed with mean vector 0 and variance-covariance $I(\beta_0)$.

Then we can define the test statistic for score test as,

$$\chi_{sc}^2 = [U(\beta_0)][I(\beta_0)]^{-1}[U(\beta_0)] \quad \text{.................(2.4)}$$

where, χ_{sc}^2 follows an asymptotic chi-square distribution with p degrees of freedom.

From the chi-square distribution table with p.d.f. we can get the value of $\chi^2_{p,\alpha y}$ at α % level of significance. If $\chi^2_{sc} > \chi^2_{p,\alpha y'}$ then we may reject the null hypothesis at α % level of significance. We may conclude that the β may not be equal to β_0.

As in linear regression model interpretation of parameters in logistic regression model is not so straightforward. Interpretation of parameters in logistic regression model can be done in terms of following two ways:

1. Interpretation in terms of logit.
2. Interpretation in terms of odds ratio.

(1) The logit transformation of logistic regression model is called the logit. The logit is defined as

$$\pi(x) = \frac{e^{\sum_{i=1}^{p} \beta_i x_i}}{1 + e^{\sum_{i=1}^{p} \beta_i x_i}} \qquad \ldots\ldots\ldots\ldots\ldots (2.6)$$

$$\log it[\pi(x)] = g(x) = \log\left[\frac{\pi(x)}{1-\pi(x)}\right]$$

$$= \beta_o + \beta_1 x_1 + \ldots\ldots\ldots\ldots + \beta_p x_p \qquad \ldots\ldots\ldots\ldots\ldots (2.7)$$

which is linear in parameters.

According to linear regression model, we can interpret the $\beta_j; (j = 1,2,\ldots\ldots\ldots, p)$ represents the rate of change in $\log it[\pi(x)]$ for one unit changes in x_j given other variables remaining constant.

(2) from logistic regression model we have,

$$\pi(x) = \frac{\exp\left(\beta_0 + \beta_1 x_1 + \ldots\ldots + \beta_i x_i + \ldots\ldots + \beta_p x_p\right)}{1 + \exp\left(\beta_0 + \beta_1 x_1 + \ldots\ldots + \beta_i x_i + \ldots\ldots + \beta_p x_p\right)}$$

where, $\pi(x)$ is the proportion of individuals with outcome being present for given x. And proportion of individuals with outcome being absent for given x is as follows,

$$1-\pi(x)=\frac{1}{1+\exp\left(\beta_0+\beta_1 x_1+\ldots\ldots+\beta_i x_i+\ldots\ldots+\beta_p x_p\right)}$$

Now, the odds of outcome being present for given x is defined as,

$$\frac{\pi(x)}{1-\pi(x)}=\exp\left(\beta_0+\beta_1 x_1+\ldots\ldots+\beta_i x_i+\ldots\ldots+\beta_p x_p\right)$$

Now we consider the situation where the independent variable is dichotomous. It is the simplest case and will provide the conceptual foundation for all the other situations. We assume that x_i takes value 0 and 1, and then the odds ratio denoted by OR is defined as the ratio of the odds for x_i=1 to the odds for x_i=0 and is given by,

$$OR=\frac{\pi(1)/[1-\pi(1)]}{\pi(0)/[1-\pi(0)]}$$

$$=\frac{\exp\left(\beta_0+\beta_1 x_1+\ldots\ldots+\beta_i\times 1+\ldots\ldots+\beta_p x_p\right)}{\exp\left(\beta_0+\beta_1 x_1+\ldots\ldots+\beta_i\times 0+\ldots\ldots+\beta_p x_p\right)}$$

$$=\exp\left(\beta_i\right)$$

Now log of the odds ratio is defined as,

$$\log OR=g(1)-g(0)=\beta_i$$

which is the logit difference or log-odds, where $g(x)$ is the logit as defined previously. Thus, we can get the estimate of the coefficients of a logistic regression model directly from $\log \widehat{OR}$ can easily be interpreted. That is the main reason why logistic regression is proved such a powerful analytic tool for epidemiologic research.

Logistic regression is not dependent on stringent assumption to be met as compared to linear regression. The fact that logistic regression analysis does not require a lot of assumptions renders it more preferable in some instances to other methods.

The following details how it differs from other techniques:

- The error terms are with a mean of zero and a variance of π(x)[1- π(x)]. (Hosmer and Lemshow, 2000)
- The conditional mean of the regression equation is greater than or equal to 0 and less than or equal to 1.
- The same principles used when conducting linear regression also apply but the difference is only that the equation will be modeling the log odds and not the actual relationship among variables.

The Pseudo R^2

There are several R^2-like statistics that can be used to measure the strength of association between the dependent and predictor variables. They are not as useful as the R^2coefficient in regression, since their interpretation is not straightforward (Hausman & McFadden, 1994). For logistic regressions, Pseudo R^2 shows the percentage of variation in the outcome variable that is explained by the logistic regression model.

Goodness-of-fit Test

In binary logistic regression, Hosmer and Lemeshow (H-L) goodness of fit test tells how closely the observed and predicted probabilities match. Also Hosmer and Lemeshow (H-L) goodness of fit statistic was another test that was used to assess the model fit. The method is similar to chi-squre goodness of fit. A very small Hosmer-Lemshow test statistic is desirable and a P-value greater than 0.05 indicates that the data model was acceptable. For multilevel logistic regressions, the test compares whether the model produces adequate accurate predictions compared to the null model (intercept only). Moreover, goodness-of-fit (measured by Pearson and Deviance statistics) also shows whether the model adequately fits the data. If the model fits well, the observed and expected cell counts will be very close, the value of the test statistic will be small and the null hypothesis cannot be rejected. It should be mentioned that if the categorical predictor has many levels or grouping, there may be cells with a small number of expected counts or many cells with zero frequencies. In that case these tests may not produce accurate or dependable goodness-of-fit test results. In such cases, it would be better to use the Pseudo R^2 (Cox & Snell, Nagelkerke, and McFadden) statistics to investigate to what extent the proposed model is an improvement over the null model and the overall classification accuracy (Agresti, 2007; Chan, 2004).

Multilevel logistic regression analysis

Multilevel logistic regression models allow one to account for the clustering of subjects within clusters of higher-level units when estimating the effect of subject and cluster characteristics on subject outcomes. A search of the PubMed database demonstrated that the use of multilevel or hierarchical regression models is increasing rapidly. However, our impression is that many analysts simply use multilevel regression models to account for the nuisance of within-cluster homogeneity that is induced by clustering. In this article, we describe a suite of analyses that can complement the fitting of multilevel logistic regression models. These ancillary analyses permit analysts to estimate the marginal or population-average effect of covariates measured at the subject and cluster level, in contrast to the within-cluster or cluster specific effects arising from the original multilevel logistic regression model (Austin PC et al., 2017).

Multilevel modeling techniques were developed to help analysts avoid erroneous conclusions from using inappropriate analysis procedures, such as using OLS regression with unadjusted standard errors when analysis nested data. In a multilevel model the associations between the dependent variable and the independent variables are expressed as regression coefficient and interpreted in the same way as OLS regression coefficients. However, in a multilevel regression analysis the co-efficients refers to specific levels (for example, students within schools) in the nested data. There are several method in Multilevel logistic regression analysis: Laplace Approximation, Numerical integration and Gauss-Hermite quadrature.

In the context of generalized linear mixed models Gauss-Hermite quadrature is a method of approximating the integral used in the calculation of the log likelihood. The quadrature location and weights for individual clusters are fixed during the optimization. Method selection allows us to specify how the analysis was performed. We use Gauss-Hermite quadrature method to construct multilevel logistic regression models from the same set of variables.

Due to this nested structure, the odds of adults experiencing the outcome of interest are not independent, because adult from the same cluster may share common exposure to community characteristics. The response variable is a binary and hence multilevel logistic regression model is a natural choice for modeling. First we introduce two level logistic regression model. For the two-level model let Y_{ij} be the binary outcome variable, coded '0' or '1' associated with level-1 unit i nested within level-2 unit j. Also let p_{ij} be the probability that the response variable equals 1, i.e. p_{ij} = Pr(Yij= 1). Here, Yij follows a Bernoulli distribution. Like logistic regression the p_{ij} is modelled using the logit link function. The two-level logistic regression model with a single explanatory variable can be written as,

$$\ln\left[\frac{p_{ij}}{1-p_{ij}}\right]=\beta_0+\beta_1 x_{ij}+u_j \qquad (1)$$

where u_j is the random effect at level 2, which is a random quantity following distribution $N(O,\sigma_u^2)$. The model (can be written as follows splitting up into two models: one for level 1 and the other for level 2.

$$\ln\left[\frac{p_{ij}}{1-p_{ij}}\right]=\beta_{0j}+\beta_1 x_{ij}+u_j \qquad \text{[model : level-1]}$$

and $$\beta_{0j}=\beta_0+u_j \qquad \text{[model : level-2]}$$

The multilevel logistic regression model can not be derived in the way simple logistic regression model is derived. This model (1) can be derived through a latent or hidden variable conceptualization (Khan MHR et al., 2011).

The proportion of variance in the dependent variable that lies between groups. ICC is useful measure for describing reliability and validity within a set of data. An intra-class correlation tells you about the correlation of the observation (cases) within a cluster. We chose to calculate the ICC based on the latent response formulation of the model as it is the approach most widely adopted in applied work. This formulation assumes a latent continuous response underlies the observed binary response and it is this latent response for which the ICC is calculated and interpreted. The higher the ICC, the more relevant cluster context is for understanding individual latent response variation (Goldstein H et al., 2002). The ICC is calculated as

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \frac{\pi^2}{3}}$$

Where $\frac{\pi^2}{3}$ denotes the variance of a standard logistic distribution. (Note that here π denotes the mathematical constant 3.1416. . ., not the probability.)

The intraclass correlation coefficient ranges from 0 to 1 and is used to estimate the degree of statistical dependency, all of the variance would be expected to lie among individuals, and the intraclass correlation coefficient would be zero or close to zero. Conversely, with highly dependent data, the largest proportion of variance would lie among groups, and so the intraclass correlation coefficient would be closer to 1 (Marlo J et al., 2016)

The MOR (Larsen K et al., 2000) is an alternative way of interpreting the magnitude of the cluster variance. The MOR translates the cluster variance estimated on the log-odds scale, to the widely used OR scale. This makes the MOR comparable with the OR of individual and cluster covariates. The MOR is defined as the median value of the distribution of ORs obtained when randomly picking two individuals with the same covariate values from two different clusters, and comparing the one from the higher risk cluster to the one from the lower risk cluster. In simple terms, the MOR can be interpreted as the median increased odds of reporting the outcome if an individual moves to another cluster with higher risk. The MOR is calculated as

$$\text{MOR}=exp\left(\sqrt{2\sigma_u^{2}}\,\Phi^{-1}(0.75)\right)$$

where $\Phi^{-1}(\cdot)$ represents the inverse cumulative standard normal distribution function. In absence of cluster variation (i.e., $\sigma_u^{2}=0$), the MOR is equal to 1.

The deviance information criterion (DIC) is widely used for Bayesian model comparison, despite the lack of a clear theoretical foundation. DIC is shown to be an approximation to a penalized loss function based on the deviance, with a penalty derived from a cross-validation argument. This approximation is valid only when the effective number of parameters in the model is much smaller than the number of independent observations. In disease mapping, a typical application of DIC, this assumption does not hold and DIC under-penalizes more complex models. Another deviance-based loss function, derived from the same decision-theoretic framework, is applied to mixture models, which have previously been considered an unsuitable application for DIC. DIC is intended as a generalization of Akaike's Information Criterion (AIC). For non-hierarchical model with little prior information, *p* approximately the true number of parameters. AIC requires counting parameters and hence any intermediate level ('random-effects') parameters need to be integrated out.

$$\text{AIC} = -2\log p(L) + 2p$$

where *L* refers to the likelihood under the fitted model and *p* is the number of parameters in the model. 2*p* is a penalty term, adjusting for the number of predictors in the model. Larger n affects -2log*L*. In hierarchical models these three techniques are essentially answering different prediction problems. AIC is a valid procedure to compare non-nested models. AIC is a better estimator of predictive accuracy, whereas BIC (see below) is a better criterion for determining process (Kuha J, 2004). Detractors contend that AIC tends to over fit the data (Bozdogan H, 1987). Note if your model or data are severely over dispersed AIC will result in biased outcomes and other model selection procedures are more appropriate.

Another widely used information criteria is the BIC. Unlike Akaike Information Criteria, BIC is derived within a Bayesian framework as an estimate of the Bayes factor for two competing models (Schwarz, 1978; Kass and Rafftery, 1995). BIC is defined as:

$$BIC = -2\log p(L) + p\log(n)$$

Superficially, BIC differs from AIC only in the second term which now depends on sample size n. Models that minimize the Bayesian Information Criteria are selected. From a Bayesian perspective, BIC is designed to find the most probable model given the data. Performance of the model selection criteria in selecting good models for the observed data is examined using simulation studies. Such a comparison is not straight forward and even its relevance could be questioned, given that the two criteria are based on different theoretical motivations and objectives. However, for application purpose, the Akaike Information Criteria and the Bayesian Information Criteria do have the same aim of identifying good models even if they differ in their exact definition of a “good model”. Comparing them is thus justified, at least to examine how each criterion performs according to recovery of the correct model or how they behave when both should prefer the same model (Acquah H, 2010).

Like the log-likelihood value on which they are largely based, these measures are functions of the target variable values, so unlike R^2 measures they cannot be used to compare models for different targets or different sets of data.

In hierarchical models, these three techniques are essentially answering different prediction problems. Suppose the two levels of our model concerned *individuals* within *clusters*. Then if we were interested in predicting results of individuals in those actual cluster, then DIC is appropriate (i.e. the random effects themselves are of interest).

The Area Under the ROC Curve is another popular summary statistic for binary classification. ROC analysis is applied to binary outcomes such as those appropriate for probit or logistic regression. After fitting a model, one can obtain predicted probabilities of a positive outcome. One chooses a value, above which the predicted probability is declared a positive and below which, a negative. A Receiver Operating Characteristic Curve (ROC) is a standard technique for summarizing classifier performance over a range of trade-offs between true positive (TP) and false positive (FP) error rates (Sweets, 1988). ROC curve is a plot of *sensitivity (the ability of the model to predict an event correctly)* versus *1-specificity* for the possible cut-off classification probability values $\pi 0$. The ROC curve is more informative than the classification table since it summarizes the predictive power for all possible $\pi 0$. The ROC (Receiver Operating Characteristic) Curve chart plots the true positive rate (TPR) on the vertical axis against the false positive rate (FPR) on the horizontal axis, as the threshold for positive classification is varied across the probability range. The true positive rate is the proportion of positive outcomes that are correctly predicted, also known as sensitivity, recall, or probability of detection. The false positive rate is the proportion of negative outcomes that are falsely predicted to be positive, also known as one minus specificity (1-specificity), fall-out, probability of false alarm, or false discovery rate (FDR).Since the predicted probabilities from a binary classification model such as a logistic regression fall in the open interval between 0 and 1, If the classification threshold is set to 1, no true or false positives would occur, so the curve begins at the (0,0) point at the lower left, and if the threshold is set to 0, all observations would be predicted to be responses, so both the true positive and false positive rates would be 1, so the curve ends at the (1,1) point at the upper right. Intermediate threshold values will produce different combinations of true and false positive rates. The diagonal line running from the lower left to the upper right of the chart represents the expected curve if classification is performed randomly, assigning positive or negative response labels to all observations with various fixed probabilities. The area under the ROC curve (AUC) is a popular summary measure of classification performance for binary classifiers. The diagonal line representing random classification divides the ROC curve space in half and corresponds to an AUC of 0.50. A model that is able to perfectly classify responses and non-responses would have an AUC of 1.00, though this is seldom seen in practice and results in non-existence of maximum likelihood estimates of one or more parameters in a logistic regression model. Typical ROC curves will have AUC values between 0.50 and 1.00. Any ROC curve with an AUC values less than 0.50 could be transformed into a curve with an AUC value above 0.50 simply by reversing the decision or group assignment rule. The fit for all models was assessed using the AUC. AUC is constructed by plotting the true positive fraction (TPF) (i.e., sensitivity) against the false positive fraction (FPF) (i.e., 1-specificity) for different binary classification thresholds of the predicted probabilities (Merlo J et al., 2016). Where a AUC of 1 is a perfect fit model and 0.5 is no better than chance. A good fit model should have a AUC >0.7 (Hosmer DW, 2000). One important advantage of the AUC measure over some other measures, such as overall classification accuracy, is that the ROC curve is based on two quantities (true and false positive rates) that are calculated from distinct parts of the observed data, actual positive and negative observations. This results in the ROC curve and the AUC measure being insensitive to changes in the relative proportions of positive and negative observations. Overall classification accuracy, on the other hand, in addition to requiring specification of a single cut point, can be highly dependent upon the relative proportions of positive and negative observations. There are dangers however in using the AUC to compare the performance of different classifiers. For example, if the ROC curves cross, it is possible for one classifier to produce a higher AUC value but to have inferior performance at critical probability thresholds that would be most useful in practice. Also, although the ROC curve and the area under it are insensitive to the relative proportions of positive and negative observations, they are dependent upon the distribution of the prediction scores. This implies the use of different assumptions of costs of misclassifications for different classifiers, making comparisons potentially akin to comparing measurements in different units.

The STATA version 13.0 has been used for computation of the results throughout the thesis. Excel has been used to create graphs and MS word has been used for writing the thesis.

2.6 Data screening

Survey screening process used to select current tobacco smokers and current smokeless tobacco users. Among the 9629 respondents aged 15 years or older who completed the survey, 2,038 were daily smokers, 195 were occasionally (less than daily) smokers, and 7,396 were former or never tobacco smokers. Thus 2,038 were current tobacco smokers [the final study subjects] (Figure 2a).

Survey screening process used to select current tobacco smokers is given below.

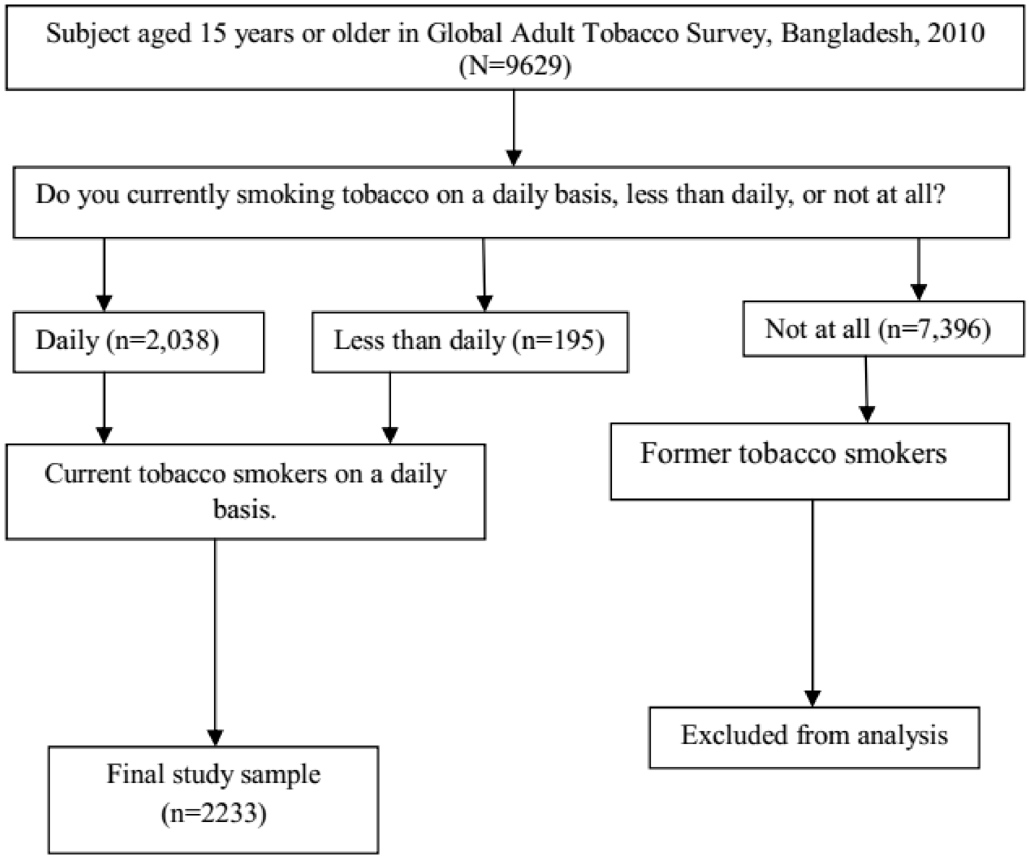

Figure 2a: Survey screening process used to select current tobacco smokers.

Among the 9629 respondents aged 15 years or older who completed the survey, 2,336 were daily smokeless tobacco users, 354 were less than daily or occasionally smokeless tobacco users, and 6,939 were former or never smokers. Thus 2,336 were current smokeless tobacco users daily [the final study subjects] (Figure 2b). Survey screening process used to select current smokeless tobacco users is given below.

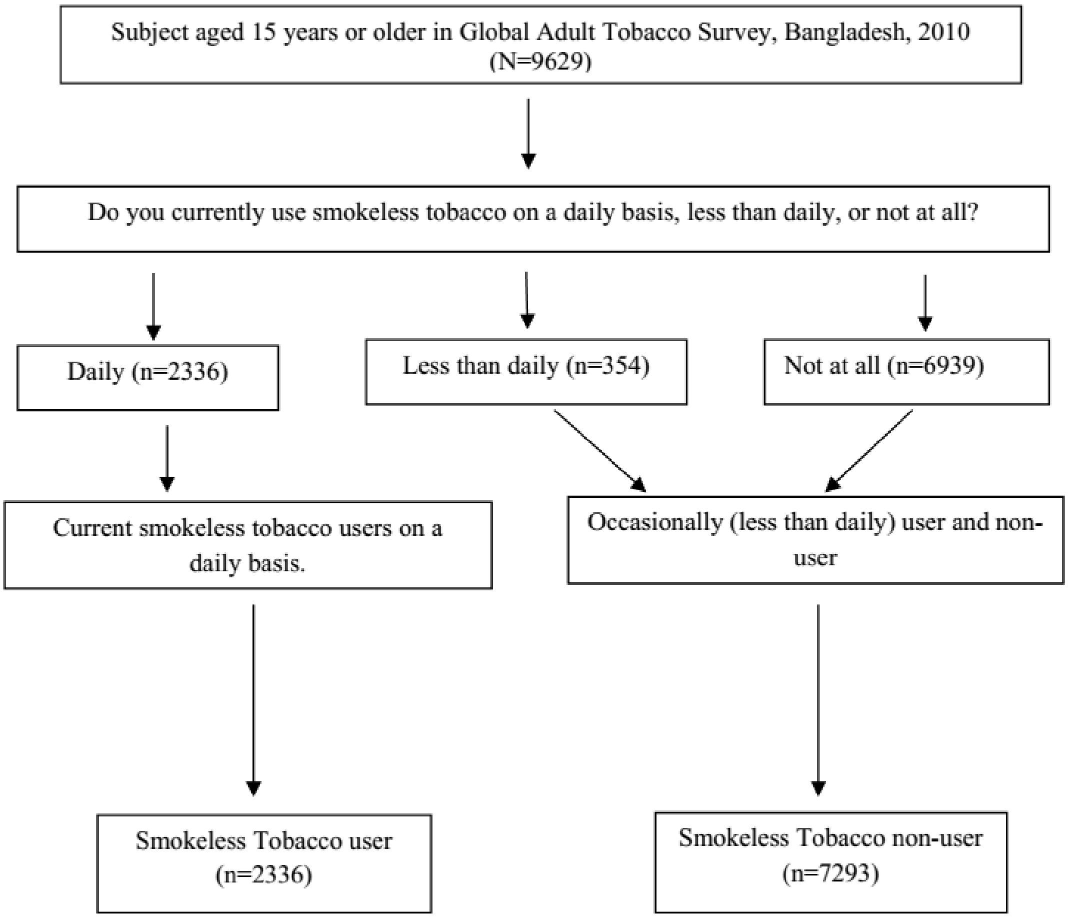

Figure 2b: Survey screening process used to select current smokeless tobacco users.

CHAPTER THREE

CRACTERISTIC OF THE STUDY SUBJECTS

3.1 Introduction

3.2 Characteristics of the study subjects

3.3 Characteristics of the study subjects to tobacco consumption

3.4 Discussion

CHAPTER THREE

CHARACTERISTICS OF THE STUDY SUBJECTS

3.1 Introduction

In the previous chapter, the data source, sampling design of the survey objectives and methodology have been discussed. Generally, it is important to know the characteristics or nature of the data before performing any advance analysis. These will be useful for discussing the results and drawing meaningful conclusions in the subsequent chapters of the report. In order to know the nature of the study subjects the frequency distribution and graphical representation could be very useful. This chapter provides information on demographic characteristics of the respondents such as residence, gender, age, educational level, occupation etc. For the purpose frequency with percentage have been reported.

This chapter also elicits information from respondents concerning:

- Tobacco smoking.
- Smokeless tobacco.

In this chapter, characteristics table and some of their corresponding graphs are presented.

3.2 Characteristics of the study subjects

Descriptive information about the sample of study was obtained from frequency of the variables considered (Table 3.2.1)

Table 3.2.1: Characteristics of study subjects (GATS, 2010)

Characteristics	**Sample size=9,629**
Residence (%)	
Urban	4,857(50.44)
Rural	4,772(49.56)
Gender (%)	
Male	4,468(46.40)
Female	5,161(53.60)
Age (yr), (%)	
≤ 24	2,073(21.53)
25-34	2,665(27.68)
35-45	2,537(26.35)
≥ 46	2,354(24.45)
Educational Level (%)	
No formal schooling	3,416(35.48)
Less than primary school completed	1,487(15.44)
Primary school completed	1,115(11.58)
Less than secondary school completed	1,937(20.12)
Secondary school completed	663(6.89)
High school completed	463(4.81)
College/University completed	273(2.84)
Post graduate degree completed	211(2.19)
Occupation (%)	
Government employee	221 (2.30)
Non-Government employee	740(7.69)
Business-small	865(8.98)
Business-large	128(1.33)
Farming (land owner & farmer)	826(8.58)
Agricultural worker	374(3.88)
Industrial worker	214(2.22)
Daily laborer	631(6.55)
Other self-employed	308(3.30)
Student	461(4.79)
homemaker/ housework	4030(41.85)
Retired	114(1.18)
unemployed, able to work	151(1.57)
unemployed, unable to work	162(1.68)
Other (Specify)	394(4.09)
Wealth Index	
Richest	1,923(19.97)
Rich	2,040(21.19)
Middle	1,732(17.99)
Poor	2,068(21.48)
Poorest	1,866(19.38)

Cont.

Table 3.2.1 Cont.

Characteristics	Sample size=9,629
Tobacco smokers	2233 (23.19)
Daily	2038 (21.17)
Less than daily	195 (2.03)
Smokeless tobacco users	2690 (27.93)
Daily	2336 (24.26)
Less than daily	354 (3.68)

Note: Tobacco smokers and smokeless tobacco users are not mutually exclusive.

Socio- economic and demographic characteristics

Socio-economic and demographic characteristics include residence, gender, age, educational level, occupation and wealth Index. Descriptive information about the sample of the study and respective graphs are given below.

Residence

The residence is an important factor in socio-economic and demographic discrepancies. This factor influence great in tobacco use. Respondents who are living in urban areas are assigned the value 1 and the value 0 is assigned to those whose residing place is in rural areas. From the characteristics Table 3.2.1, we have observed that 50.44% adults of 15 years or above are urban and 49.56% are rural, which are approximately equal.

Residence categories are shown in the pie chart as follows

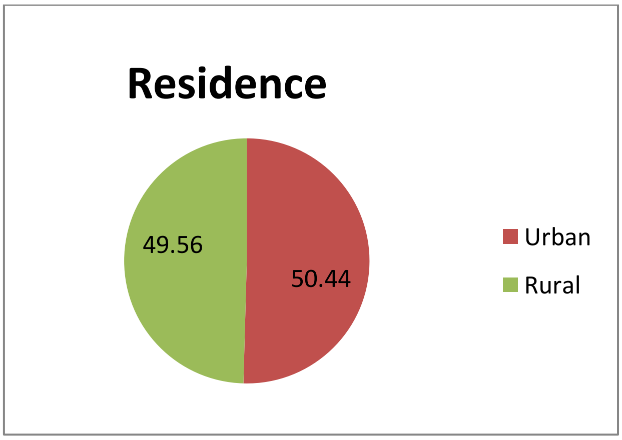

Figure: 3.2.1 Pie chart of residence

Gender

Gender is the most important factor for tobacco use and tobacco related death. Gender interacts with the social, economic and biological determinants. Also gender is a measure of both biological and social differences. The value is assigned for male and the value is assigned for female. From the characteristics Table 3.2.2, we have observed that 46.40% are male and 53.60% are female. So among the respondents female are more than male.

Gender categories are shown in the pie chart as follows

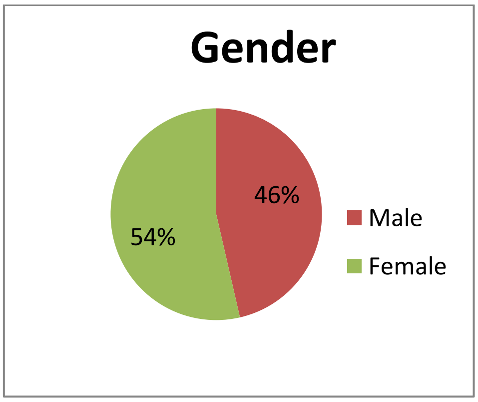

Figure 3.2.2: Pie chart of gender

Age

The age of respondents varied, with the greatest proportion (27.68%) in the range of 25-34 years old. 21.53% of the respondents were in the youngest age group of $\leq$ 24 years. Approximately 26% of the adults in Bangladesh were in the age group of 35-45 years. About 24% of the respondents were in the age group of 46 years and above.

Age categories are shown in the pie chart as follows

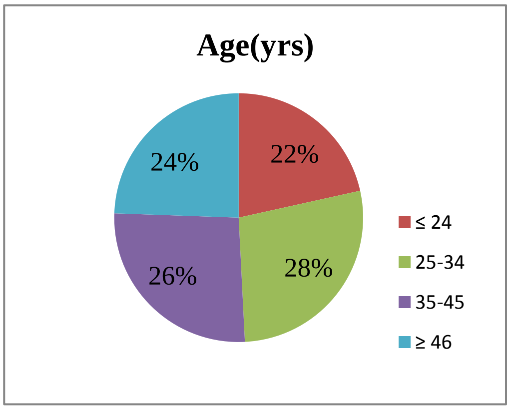

Figure 3.2.3: Pie chart of Age(yrs)

Educational Level

Education is the most important factor for showing the pattern of tobacco use. From the characteristics Table 3.2.1, we have observed that educational level is highest among respondent with a no formal schooling is 35.48%, lowest among those with a post graduate degree completed is 2.19%, followed by less than primary school completed 15.44%, primary school completed 11.58%, less than secondary school completed 20.12%, secondary school completed 6.89%, high school completed 4.81%, Only 2.84% of the adults in Bangladesh had completed higher education (college or university).

Education categories are shown in the bar chart as follows

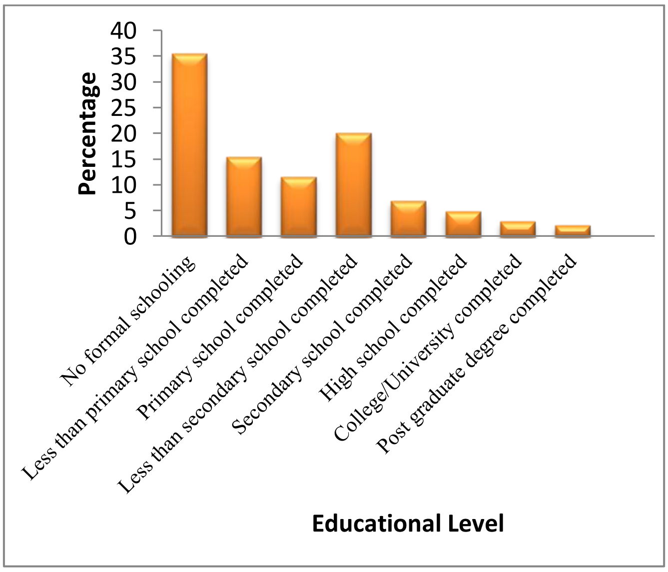

Figure 3.2.4: Bar chart of Educational Level

Occupation

Occupation is also an important factor because use of tobacco may depend on different occupation. Occupation is coded into fourteen categories. From the characteristics Table 3.2.1, we have observed that by occupation, 41.89% is highest among homemaker/housework and lowest among retired people is 1.17%. Government employee is 2.30%, followed by non-government employee is 7.69%, followed by business-large 8.99%, farming 1.33%, agricultural work 8.59%, industrial worker 2.22%, agricultural worker 3.89%, daily laborer 6.56%, student 4.78%, unemployed, able to work 1.53%; unemployed, unable to work 1.68% and other 4.03%.

Occupation categories are shown in the bar chart as follows

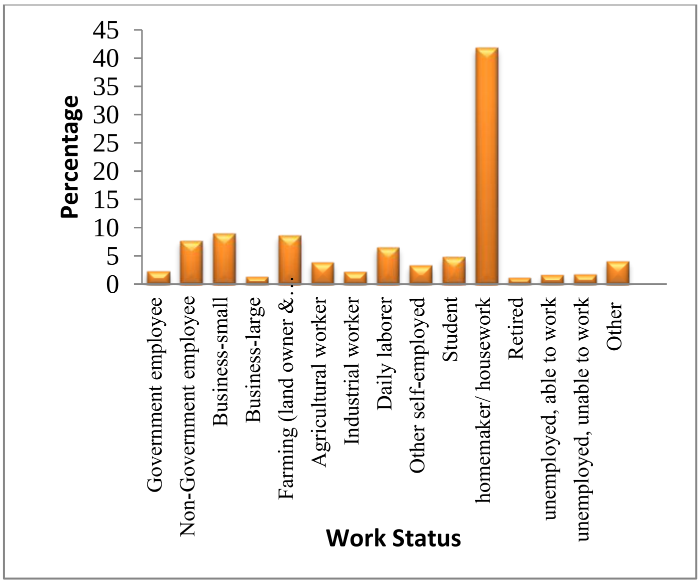

Figure: 3.2.5: Bar chart of work status

Wealth Index

About 21.48% of the adults and 19.38% of the respondents in Bangladesh were from the lowest two quintiles of wealth index (poor category) whereas, 19.97% and 21.19% in Bangladesh were from the highest two quintiles (rich category). The 17.99% of all respondents fell into the middle income category.

Wealth Index categories are shown in the bar chart as follows

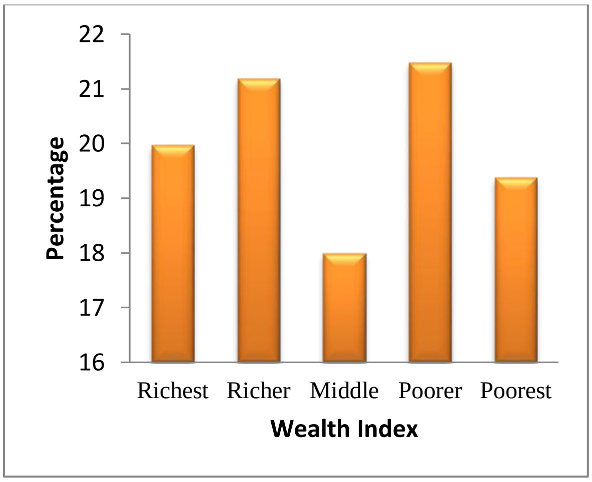

Figure 3.2.6: Bar chart of wealth index

Among the respondents 23.19% were tobacco smokers (21.17% daily smokers, 2.03% less than daily). Also 27.93% of the respondents were smokeless tobacco users (24.26% daily users, 3.68% less than daily users). (Table 3.2.7)

Pattern of tobacco using are shown in figure 3.2.7 below.

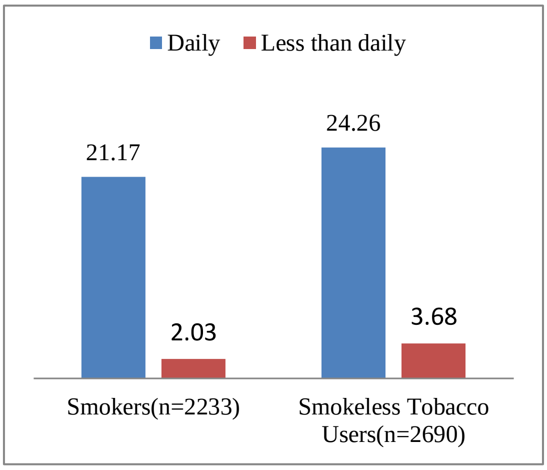

Figure 3.2.7: Pattern of tobacco using

3.3 Characteristics of the study subjects to tobacco consumption

A tobacco user was defined as someone who currently smokes (cigarettes, bidis, hookah, or other smoked products) and/or uses smokeless tobacco products at least once a month. The Tobacco smokers were defined as those who reported that they smoke currently on daily basis or less then daily basis and

Any individual who did not meet these criteria was classified as a non user of tobacco. Tobacco users were further categorized according to the products they currently use. Within the dataset, it was possible to divide respondents into four categories: smokers, smokeless tobacco users, mixed users (people who both smoke and use smokeless tobacco), and non users (those who do not currently use any tobacco products). TC included smoked and smokeless tobacco products. The problem of adult TC seemed more serious in Bangladesh and more among the male **(**Table 3.3.1**).**

Table 3.3.1: Characteristics of Study subjects to Tobacco Consumption (GATS, 2010)

Characteristics	Tobacco Users (n=4242)	Tobacco Smokers (n=2233)	Smokeless Tobacco Users (n=2690)
Residence (%)			
Urban	1,939(45.71)	964(41.30)	1,165(43.31)
Rural	2,303(54.29)	1074(52.74)	1,525(56.69)
Gender (%)			
Male	2,762(65.11)	1,972(96.76)	1,251(46.51)
Female	1,480(34.89)	66(3.24)	1,439(53.49)
Age (yr), (%)			
≤ 24	339(7.99)	197(9.67)	140(5.20)
25-34	949(22.37)	517(25.37)	517(19.22)
35-45	1,373(32.37)	678(33.27)	892(33.16)
≥ 46	1,581(37.27)	646(31.70)	1,141(42.42)
Educational Level (%)			
No formal schooling	2,128(50.50)	958(47.01)	1,469(55.18)
Less than primary school completed	706(16.75)	372(18.25)	430(16.15)
Primary school completed	424(10.06)	184(9.03)	286(10.74)
Less than secondary school completed	570(13.53)	315(15.46)	300(11.27)
Secondary school completed	161(3.82)	88(4.32)	82(3.08)
High school completed	93(2.21)	51(2.50)	40(1.50)
College completed or higher	79(3.13)	70(3.43)	55(2.07)

Table 3.3.1 Con.

Occupation (%)			
Employment (Govt/Non-Govt.)	381(8.98)	225(11.04)	180 (6.69)
Business (small/large)	621(14.64)	440(21.59)	280(10.41)
Farming (land owner & farmer)	581(13.70)	411(20.17)	293(10.89)
Agri or industrial worker/Daily labour/ Other self employed	1,023(24.12)	702(34.45)	507(18.85)
homemaker/ housework	1,154(27.20)	42(2.06)	1,127(41.90)
Retired and unemployed (able/unable to work)	224(5.28)	83(4.07)	148(5.50)
Student/ Others	258(6.08)	135(6.62)	155(5.76)
Wealth Index			
Poorest	1,042(24.56)	263(12.90)	707(26.28)
Poor	1,043(24.59)	392(19.23)	660(24.54)
Middle	773(18.22)	375(18.40)	491(18.25
Rich	826(19.47)	516(25.32)	514(19.11)
Richest	558(13.15)	492(24.14)	318(11.82)

Note: TC=Tobacco Consumption, Tobacco users, n=4242, i.e. 44.05% of the respondents are tobacco users. Tobacco smokers and smokeless tobacco users are not mutually exclusive.

The prevalence of tobacco use was 54.29% and 45.71% in the rural and urban samples, respectively. Prevalence of smoking in the rural respondents was 52.74% and 41.30% in the urban respondents. The prevalence of smokeless tobacco use in rural respondents (56.69%), was higher than in the male sample (43.31%). The distribution of tobacco consumption by male and female is shown in figure 3.3.1.

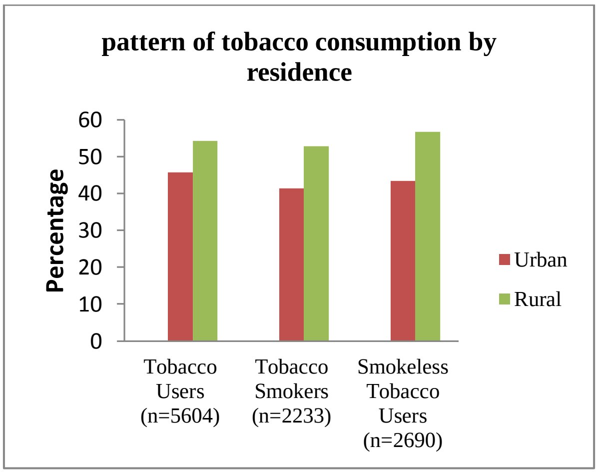

Figure 3.3.1: Distribution of tobacco consumption by urban and rural respondents

The prevalence of tobacco use was 65.11% and 34.89% in the male and female samples, respectively. Prevalence of smoking in the male sample was 96.76% and 3.24% in the women sample. The prevalence of smokeless tobacco use in the female sample, (53.49%), was marginally higher than in the male sample (46.51%).

The distribution of tobacco consumption by male and female is shown in figure 3.3.2 below.

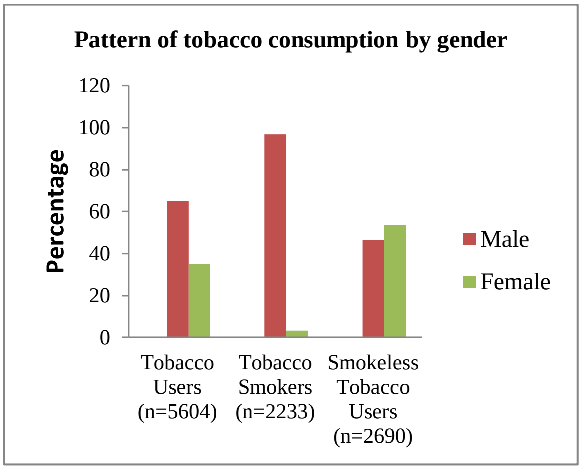

Figure 3.3.2: Distribution of tobacco consumption by male and female

The age of respondents varied, with the greatest proportion in the range of 46 years and above old 37.27% are of tobacco users Very few tobacco users were in the youngest age group of ≤ 24 years (7.99%). With increasing age, the prevalence of tobacco smoking and smokeless tobacco use also increased reaching a peak in 46 years and above age group (31.70% vs 42.42%) and declining thereafter. Tobacco consumption to different age group is shown in figure 3.3.3 below.

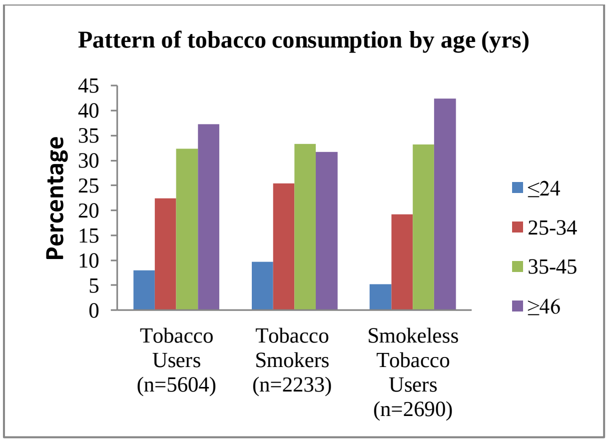

Figure 3.3.3: Tobacco Consumption by age group

50.71% of the tobacco users had no formal schooling. The majority of tobacco users were homemaker/housework (27.20%). There was a general declining trend with increasing education. Tobacco smoking was highest in no formal schooling (47.01%) followed by in less than primary school completed (18.25%), primary school completed (9.03%), less than secondary school completed (15.46%) , secondary school completed (4.32%) and high school completed (2.50%). Also smokeless tobacco use was highest in no formal schooling (55.18%), less than primary school completed (16.15%), primary school completed (10.74%), less than secondary school completed (11.27%), secondary school completed (3.80%%) and high school completed (1.50%).

Tobacco consumption by educational level is shown in figure 3.3.4 below.

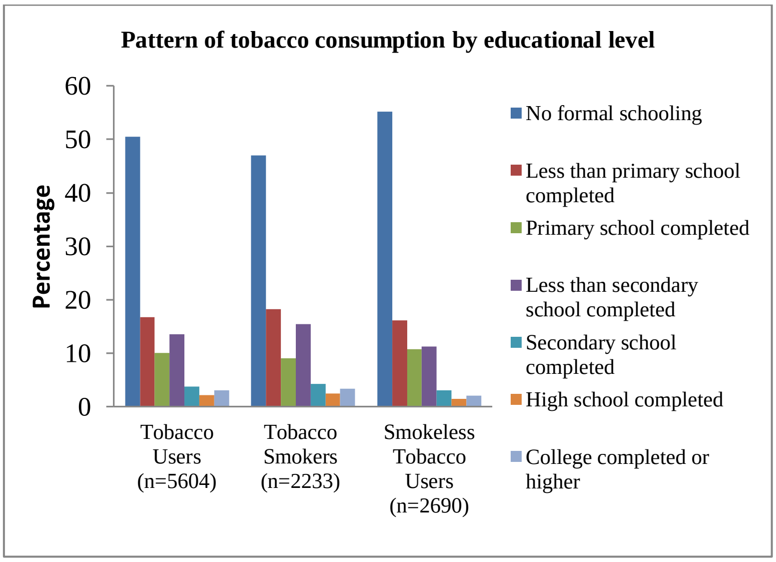

Figure 3.3.4: Tobacco consumption by educational level

Tobacco smoking was more prevalent among Agri or industrial worker/Daily labour/ Other self employed (34.45%) adults than in the other occupational categories. By occupation, prevalence of smokeless tobacco use among adults was highest among the homemakers (41.90%) (Table 3.2.1). Tobacco consumption by work status is shown in figure 3.3.5 below.

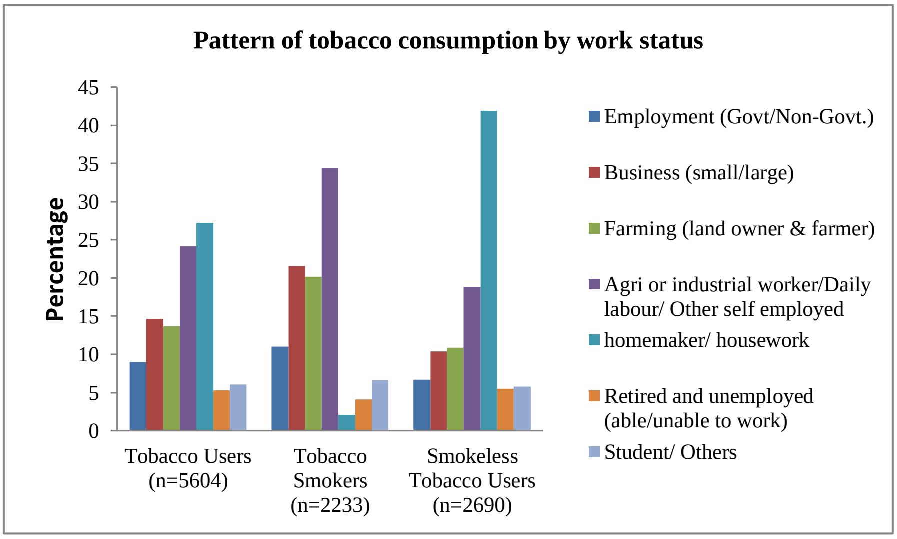

Figure 3.3.5: Tobacco consumption by occupational status

In addition, the majority of tobacco users fell into the poor category (24.59%) and the majority of tobacco smokers (19.23%) and smokeless tobacco users (26.28%) fell into poor and poorest category (25.34%) respectively. Tobacco consumption by economic status is shown in figure 3.3.6 below.

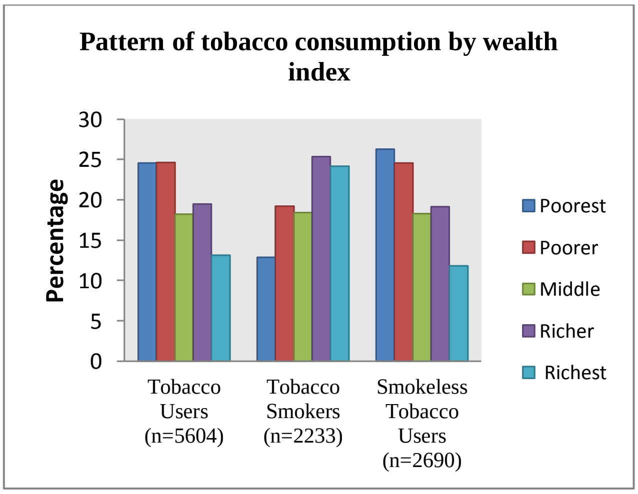

Figure 3.3.6: Tobacco consumption by economical status

3.4 Discussion

In this survey we observed that male and female respondents were approximately equal. Also respondent from rural and urban were approximately same. About 44.01% of the adults were consuming some form of tobacco products (either smoked or smokeless). Male smoked more than female. However, smokeless tobacco products were relatively more popular among females among Bangladesh. By education, the respondent with no formal schooling smoked most. By occupation, most of the smokeless tobacco users were homemaker/housework and most of the smokers were agricultural or daily labourer/industrial worker. Those who were poorer in Bangladesh consumed tobacco most. The respondents who used tobacco usually they were daily users.

CHAPTER FOUR

PATTERN AND PREVALENCE OF TOBACCO USE IN BANGLADESH

CHAPTER FOUR

PATTERN AND PREVALENCE OF TOBACCO USE IN BANGLADESH

4.1 Introduction

In the previous chapter, the information on demographic characteristics of the study subjects has been provided through frequency distribution, percentage and graphical representation. In this chapter distribution of tobacco products have been performed. The nature of the respondents i.e. their usual place of residence, age, level of education, economic status, use of tobacco have been observed. In this chapter, a comparison of socio-economic and demographic characteristics of the study subjects to current tobacco user has been performed. To compare variables chi-square test (Pearson Chi-square or Likelihood Ratio Chi-square whichever applicable) has been used, and prevalence with 95% confidence interval has been reported for individual variable. These tests have been performed at 5% level of significance. Age adjusted and unadjusted prevalence has been reported with prevalence and 95% confidence interval, too.

4.2 Tobacco Smoking

4.2.1 "Current" pattern of Tobacco Smoking in Bangladesh (GATS, 2010).

In this section, current pattern of tobacco smoking have been summarized in Table 4.2.1. The problem of adult tobacco smoking seemed more serious among the male (41.24% from urban areas and 47.16% from rural areas).

Table 4.2.1: Prevalence of "Current" Tobacco Smoking

"Current" tobacco smokers	Urban(n=4857)			Rural(n=4772)			Total(n=9629)		
	Male	Female	P-value	Male	Female	P-value	Male	Female	P-value
Daily	41.24	0.86	<0.001	47.16	1.70	<0.001	44.14	1.28	<0.001
Less than daily	4.20	0.16	<0.001	4.08	0.23	<0.001	4.14	0.19	<0.001
Not at all	54.55	98.99	<0.001	48.76	98.07	<0.001	51.72	98.52	<0.001

Note: Less than daily indicates occasionally. P-values are obtained from *Z*-test for proportion.

In Bangladesh tobacco smoking in females from rural areas (1.70%) and females from urban areas (0.86%) was very low. Smoking rates (daily) are higher among men than among women (44.14% vs 1.28% in Bangladesh). Prevalence of current tobacco smoking is shown in the Figure 4.2.1 below

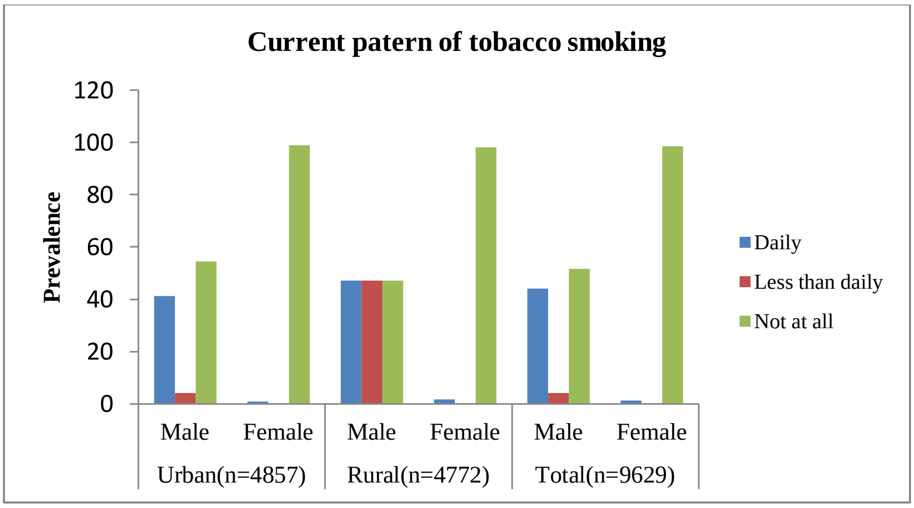

Figure 4.2.1: Current pattern of tobacco smoking by residence and gender.

Also among the adults 4.14% of male and 0.19% of female are less than daily smoker. Finally, the majority of non users of tobacco smoking in the sample (98.52%) were female.

4.2.2 Comparison of socio-demographic and economic characteristics of study subjects to "current" tobacco smoking.

To compare variables, Bivariate cross tabulations of the response variable on the independent variables were generated using Chi-square test (Pearson Chi-square or Likelihood Ratio Chi-square whichever applicable).

Table 4.2.2: Tobacco smoking daily by socio-economic and demographic factors (n=9629)

Socio-economic and demographic status	"Current" tobacco smokers		P-value*
	Users (n%)	Non Users (n%)	
Residence (%)			
Urban	964(41.30)	3893(51.28)	0.001
Rural	1074(52.74)	3698(48.72)	
Gender (%)			
Male	1,972(96.76)	2,496(32.88)	<0.001
Female	66(3.24)	5,095(67.12)	
Age (yrs) (%)			
≤ 24	197(9.67)	1,876(24.71)	<0.001
25-34	517(25.37)	2,148(28.30)	
35-45	678(33.27)	1,859(24.49)	
≥46	646(31.70)	1,708(22.50)	
Educational Level (%)			
College completed or higher	70(3.43)	414(5.50)	<0.001
High school completed	51(2.50)	412(5.47)	
Secondary school completed	88(4.32)	575(7.64)	
Less than secondary school completed	315(15.46)	1,622(21.55)	
Primary school completed	184(9.03)	931(12.37)	
Less than primary school completed	372(18.25)	1,115(14.81)	
No formal schooling	958(47.01)	2,458(32.66)	
Occupation (%)			
Employee (Govt. Non-Govt.)	225(11.04)	736(9.70)	<0.001
Business (small/large)	440(21.59)	553(7.28)	
Farming (land owner and farmer)	411(20.17)	415(5.47)	
Agri / industrial worker/daily labour/ other self employed	702(34.45)	835(11.00)	
homemaker/housework	42(2.06)	3,988(52.54)	
Retired and unemployed (able/unable	83(4.07)	344(4.53)	
Student/Others	135(6.62)	720(9.48)	

Table 4.2.2 Cont.

Socio-economic and demographic status	"Current" tobacco smokers		P-value*
	Users (n%)	Non Users (n%)	
Wealth Index			
Richest	263(12.90)	1660(21.87)	<0.001
Richer	392(19.23)	1648(21.71)	
Middle	375(18.40)	1357(17.88)	
Poorer	516(25.32)	1552(20.45)	
Poorest	492(24.14)	1374(18.10)	

*P-values are obtained from Chi-square test (Pearson Chi-square or Likelihood Ratio whichever applicable). 95% CIs are obtained from p±1.96 se (p). – indicates N/A due to no observation. Total sample size=9629.

These tests have been performed at 5% level of significance. The results showed that gender, residence, age (yrs), educational level, work status, wealth index were significantly ($p<0.001$) associated with tobacco smoking (Table 4.2.2). The majority of tobacco smokers lived in rural areas (52.74%) as opposed to urban areas (41.30%) whereas 51.28% urban adults and 48.72% are non users. Figure 4.2.2 shows the current tobacco smoking by residence.

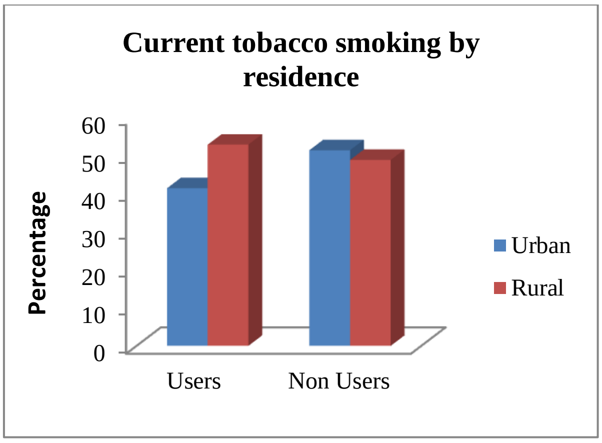

Figure 4.2.2: Current tobacco smoking by Residence.

Again, the majority of tobacco smokers (96.76%) were male while the majority of non users (67.12%) were female. Only 3.24% female were tobacco smokers.

Figure 4.2.3 shows the current tobacco smoking by Gender

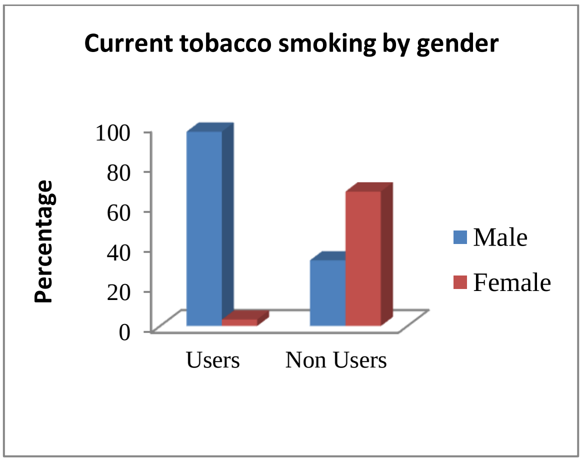

Figure 4.2.3: Current tobacco smoking by Gender

In Bangladesh, the higher the education level, the lower the likelihood to be a tobacco smokers. Respondents with no formal schooling (47.01%) had a higher tendency to tobacco smoking. However, for adults who completed higher education, only 3.43% were current tobacco smoking. Similarly the prevalence of tobacco smokers was 18.25%for less than primary school completed and 15.46% for less than secondary school completed. Figure 4.2.4 shows the current tobacco smoking by educational level.

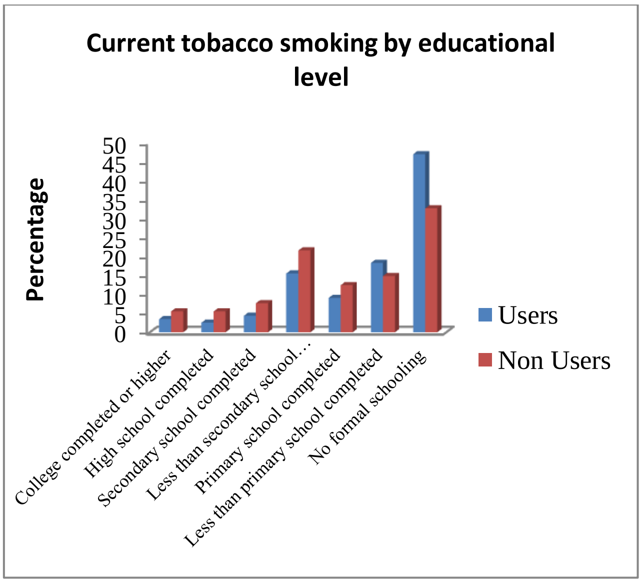

Figure 4.2.4: Current tobacco smoking by educational level

Among the adults aged 35-45, the prevalence of tobacco smoking was 33.27%. The prevalence for adults aged 15-24, 25-34 and 46 and above years old had the prevalence which were 9.67%, 25.37% and 31.70% respectively.

Figure 4.2.5 shows the current tobacco smoking by age (yrs).

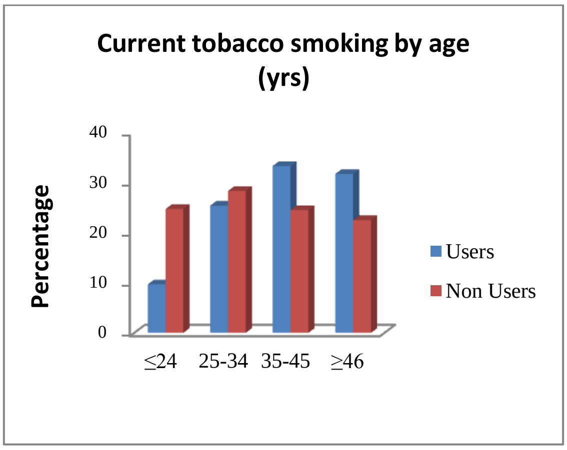

Figure 4.2.5 shows the current tobacco smoking by age (yrs)

The tobacco smoking prevalence rate (25.32%) was also higher among those who were from poorer group whereas 24.14%, 18.40%, 19.23% and 12.90% were respectively from poorest, middle, richer and richest group. Figure 4.2.6 shows the current tobacco smoking by wealth index.

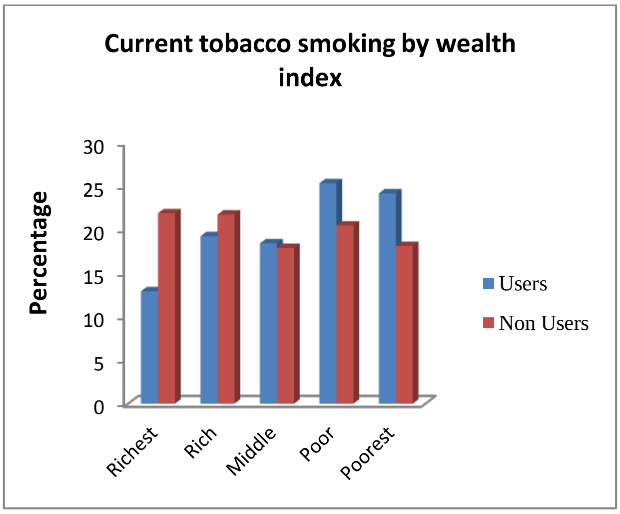

Figure 4.2.6 Current tobacco smoking by wealth index

Tobacco smoking was more prevalent among Agricultural or industrial worker/daily labour/ other self employed (34.45%) adults than in the other occupational categories; for instance, 21.59% for business (small/large), 20.17% farming (land owner/and farmer)(Table 4.3.2). Figure 4.2.7 shows the current tobacco smoking by work status.

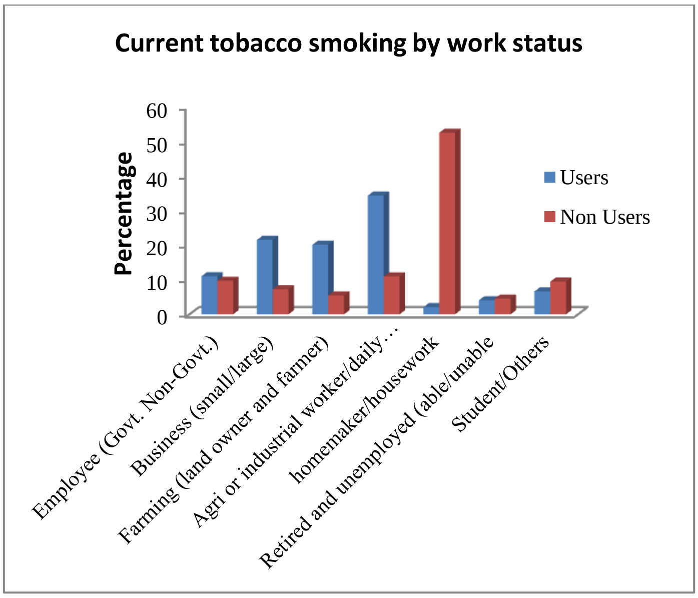

Figure 4.2.7 Current tobacco smoking by work status

4.2.3 Current pattern of smoking tobacco product used among the users in Bangladesh stratified by area of residence and gender

The distribution of tobacco products used by residence and gender is shown in Table 4.2.3 below

Table 4.2.3: Pattern of current smoked tobacco products (GATS, 2010) in Bangladesh

Categories of smoked tobacco product*		Urban (n=964)			Rural (n=1074)			Total(n=2038)		
		Male % (95% CI)	Female % (95% CI)	p-value	Male % (95% CI)	Female % (95% CI)	p-value	Male % (95% CI)	Female % (95% CI)	p-value
Manufactured cigarettes	daily	85.56 (83.32-87.81)	13.64 (0.00-27.98)	<0.001	48.45 (45.39-51.49)	11.36 (1.99-20.74)	<0.001	66.18 (64.09-68.26)	12.12 (4.25-19.99)	<0.001
	Less than daily	0.21 (0.00-0.50)	0	0.829	1.94 (1.09-2.78)	0	0.351	1.12 (0.65-1.58)	0	0.388
	Not at all	14.23 (11.99-16.46)	86.36 (72.02-100.70)	<0.001	49.61 (46.56-52.67)	88.64 (79.26-98.01)	<0.001	32.71 (30.64-34.78)	87.88 (80.00-95.75)	<0.001
Handrolled cigarettes	Daily	3.29 (2.15-4.43)	0	0.387	0.87 (0.31-1.44)	4.55 (0.00-10.70)	0.018	2.03 (1.41-2.65)	3.03 (0.00-7.17)	0.573
	Less than daily	1.06 (0.41-1.72)	0	0.627	0	0	-	0.50 (0.19-0.82)	0	0.562
	Not at all	95.65 (94.34-96.95)	1 (-)	0.317	99.13 (98.56-99.69)	95.45 89.29-101.60)	0.018	97.46 (96.77-98.16)	96.97 (92.83-101.1)	0.802
pipes full of tobacco	Daily	3.08 (1.98-4.18)	0	0.403	0.87 (0.31-1.44)	6.82 (0.00-14.27)	<0.001	1.93 (1.32-2.53)	4.55 (0.00-9.57)	0.136
	Less than daily	0.96 (0.33-1.58)	0	0.645	0	0	-	0.46 (0.16-0.75)	0	0.582
	Not at all	95.96 (94.71-97.22)	1 (-)	0.336	99.13 (98.56-99.69)	93.18 (85.73-100.63)	<0.001	97.62 (96.94-98.29)	95.45 (90.43-100.48)	0.264

Table 3.2.3 Cont.

Categories of smoked tobacco product*		Urban (n=964)			Rural (n=1074)			Total(n=2038)		
		Male % (95% CI)	**Female % (95% CI)**	**p-value**	**Male % (95% CI)**	**Female % (95% CI)**	**p-value**	**Male % (95% CI)**	**Female % (95% CI)**	**p-value**
cigars, cheroots, or cigarillos	Daily	3.39 (2.24-4.55)	0	0.379	0.87 (0.31-1.44)	0	0.534	2.08 (1.45-2.71)	4.55 (0.00-9.57)	0.237
	Less than daily	0.96 (0.33-1.58)	0	0.645	0	0	-	0.46 (0.16-0.75)	0	0.582
	Not at all	95.65 (94.34-96.95)	1 (1-1)	0.317	99.13 (98.56-99.69)	1 (1-1)	0.534	97.46 (96.77-98.16)	1 (1-1)	0.190
water pipe (hukkah)	Daily	3.93 (2.69-5.17)	4.55 (0.00-13.25)	0.883	2.72 (1.73-3.71)	13.64 (3.50-23.78)	<0.001	3.29 (2.51-4.08)	10.61 (3.18-18.03)	0.002
	Less than daily	1.27 (0.56-1.99)	0	0.594	0.09 (0.00-0.29)	0	0.836	0.65 (0.30-1.02)	0	0.508
	Not at all	94.79 (93.38-96.22)	95.45 (86.75-104.16)	0.891	97.18 (96.17-98.19)	86.36 (76.22-96.50)	<0.001	96.04 (95.18-96.90)	89.39 (81.97-96.82)	0.008
Bidis	Daily	23.04 (20.35-25.72)	86.36 (72.02-1.01)	<0.001	60.87 (57.89-63.85)	70.45 (56.97-83.94)	0.201	42.79 (40.62-44.98)	75.76 (65.42-86.09)	<0.001
	Less than daily	0.63 (0.13-1.14)	0	0.707	0.45 (0.06-0.91)	0	0.643	0.56 (0.23-0.89)	0	0.543
	Not at all	76.33 (73.61-79.04)	13.64 (0.00-27.98)	<0.001	38.64 (35.67-41.61)	29.55 (16.06-43.03)	0.224	56.64 (54.46-58.83)	24.24 (13.90-34.58)	<0.001
Dhaba (Like Water Pipe - used in Hill, Marijuana, Zarda	Daily	3.39 (2.24-4.55)	0	0.379	1.65 (0.87-2.43)	18.18 (6.79-29.58)	<0.001	2.48 (1.79-3.17)	12.12 (4.25-19.99)	<0.001
	Less than daily	0.74 (0.19-1.29)	0	0.685	0	0	-	0.35 (0.09-0.62)	0	0.628
	Not at all	95.86 (94.59-97.13)	1 (-)	0.33	98.35 (97.57-99.13)	81.82 (70.42-93.21)	<0.001	97.61 (96.43-97.89)	87.88 (80.00-95.75)	<0.001

Any products include: Betel quid with supari, Betel quid with zarda, Pan Masala, Sada pata, Sada pata with dry Banana leaf mixture, Sada pata with neam leaf mixture, Tobacco powder, use Tobacco leaf for brushing teeth, Zarda, Zarda with Supari. '-' no observations ; Respondents who smoked products daily.

Male respondents of 44.14% were tobacco smokers and female respondents of 1.28% were tobacco smokers only (Table 4.2.1). The most smoked tobacco product was manufactured cigarettes (66.18%) by male compared to hand-rolled cigarettes (2.03%), pipes full of tobacco (3.08), cigars, cheroots, or cigarillos (3.39%), water pipe (hukkha) (3.39%), Bidis (23.04) or other products including dhaba (3.39%) whereas the majority of female smokers smoked bidis (75.76%), compared to manufactured cigarettes (12.12%), hand-rolled cigarettes (3.03%), pipes full of tobacco (4.55%), cigars, cheroots, or cigarillos (4.55%), water pipe (hukkha) (10.61%) or other products including dhaba (12.12%). The distribution of tobacco products used by men and women is shown in Figure 4.2.8 below.

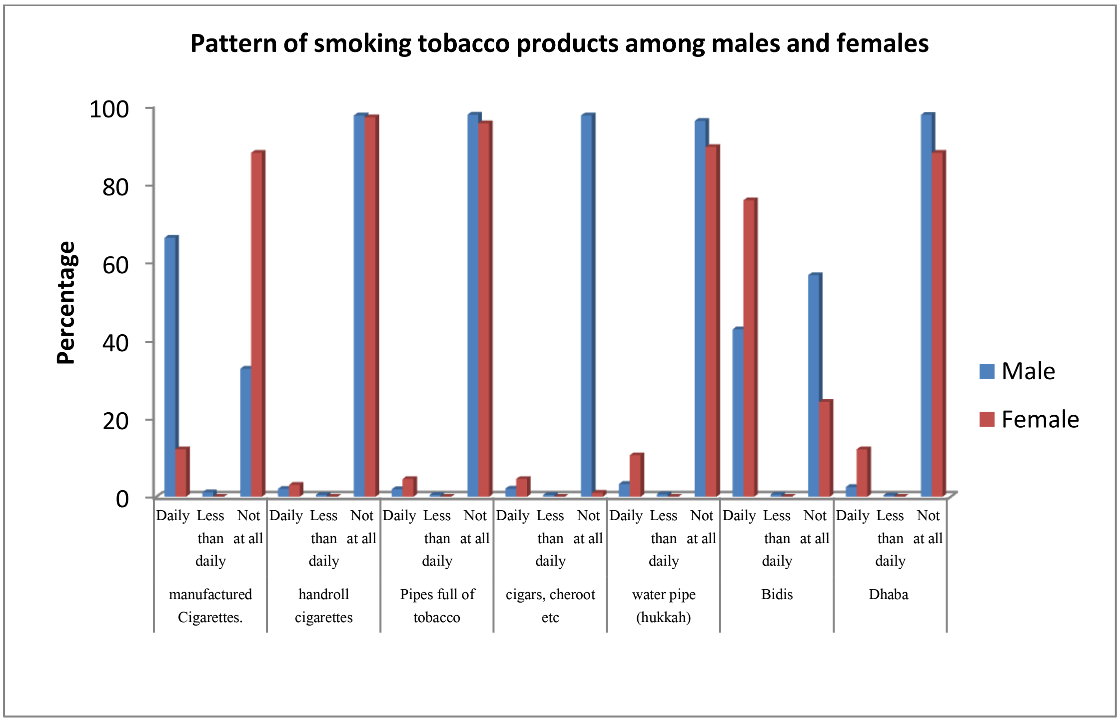

Figure 4.2.8: Pattern of using smoked tobacco products among males and females

The pattern of using smoked tobacco products for urban males versus urban females – 85.56% male used manufactured cigarettes compared to hand-rolled cigarettes 3.29%), pipes full of tobacco (3.08), cigars, cheroots, or cigarillos (3.39%), water pipe (hukkha) (3.39%), Bidis (23.04) or other products including dhaba (3.39%) whereas female smokers smoked manufactured cigarettes (13.64%), hand-rolled cigarettes (0%), pipes full of tobacco (0%), cigars, cheroots, or cigarillos (0%), water pipe (hukkha) (4.55%), Bidis (86.36%) or other products including dhaba (0%). The distribution of tobacco products by male and female are shown in the Figure 4.2.9 below.

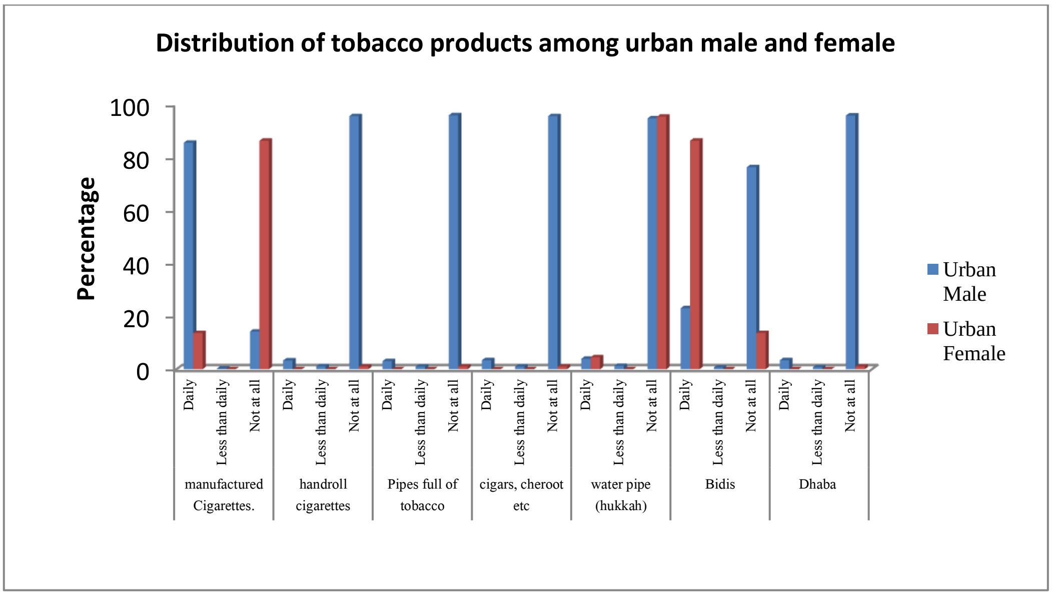

Figure 4.2.9: Distribution of tobacco products used by urban male and female

the pattern of using smoked tobacco products (daily) for rural males versus rural females – 48.45% male used manufactured cigarettes compared to hand-rolled cigarettes 0.87%), pipes full of tobacco (0.87%), cigars, cheroots, or cigarillos (0.87%), water pipe (hukkha) (2.72%), Bidis (60.87%) or other products including dhaba (1.65%) whereas female smokers smoked manufactured cigarettes (11.36%), hand-rolled cigarettes (4.55%), pipes full of tobacco (6.82%), cigars, cheroots, or cigarillos (0%), water pipe (hukkha) (13.64%), Bidis (70.45%) or other products including dhaba (18.18%).

The distribution of tobacco products by rural male and rural female are shown in the Figure 4.2.10 below.

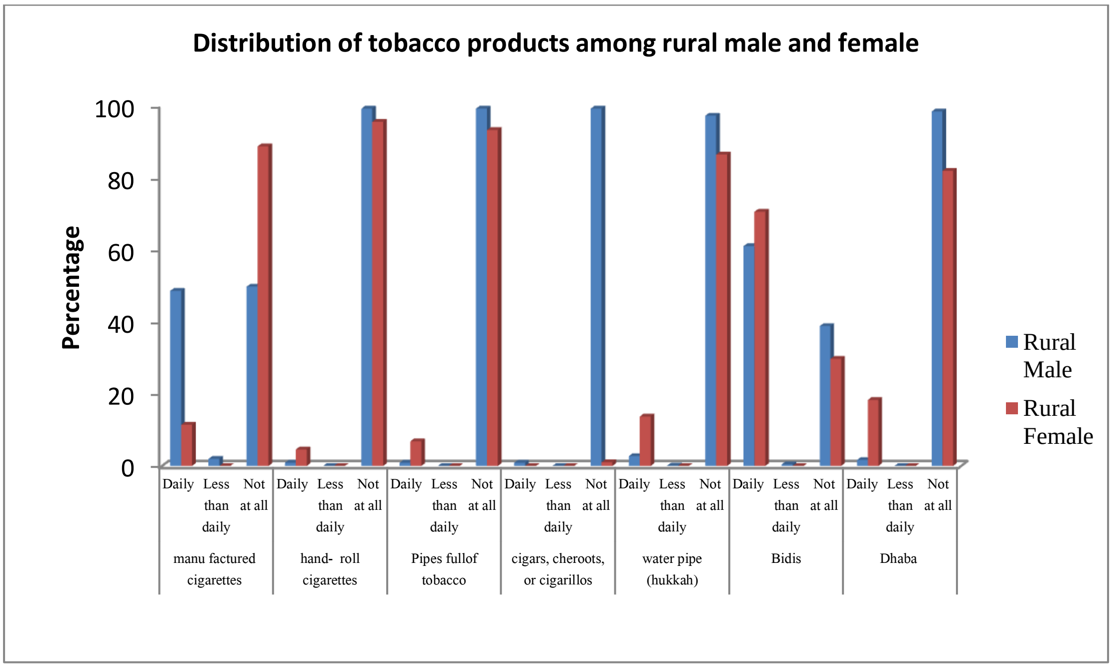

Figure 4.2.10: Distribution of tobacco products rural male and female

4.2.4 Age adjusted and unadjusted prevalence of tobacco smoking

Age adjusted and unadjusted prevalence of tobacco smoking among adults by selected socio-demographic characteristics is presented in Table 5.2.1. It was found that age adjusted and unadjusted prevalence of tobacco smoking differs only for work status which is expected. Prevalence of tobacco smoking among the urban adults was 39.92 % and for rural adults was 48.26 %. Approximately same result was observed when age was adjusted (41.06% for urban and 47.11% for rural).

Table 4.2.4: Age adjusted and unadjusted prevalence of tobacco smoking

Socio-economic and demographic status	Prevalence (95%CI) Unadjusted	Prevalence (95% CI) Age Adjusted
Residence		
Urban	39.92 (38.54 - 41.30)	41.06 (39.77- 42.36)
Rural	48.26 (46.84-49.68)	47.11 (45.78-48.45)
Gender		
Male	61.82 (60.39-63.24)	59.91 (58.53-61.28)
Female	28.67 (27.44-29.91)	30.27 (29.04-31.46)
Educational Level		
College completed or higher	27.27 (23.31-31.24)	26.68 (23.03-30.34)
High school completed	20.09 (16.44-23.74)	23.60 (19.73-27.47)
Secondary school completed	24.28 (21.02-27.55)	29.26 (25.79-32.74)
Less than secondary school completed	29.43 (27.40-31.46)	35.20 (33.07-37.34)
Primary school completed	38.03 (35.18-40.88)	41.79 (39.01-44.58)
Less than primary school completed	47.48 (44.94-50.02)	48.15 (45.74-50.56)
No formal schooling	62.29 (60.67-63, 92)	55.92 (54.23-57.60)
Work Status		
Employee (Govt. Non-Govt.)	39.64 (36.55-42.74)	41.63 (38.80-44.47)
Business (small/large)	62.54 (59.53-65.55)	59.75 (56.89-62.60)
Farming (land owner and farmer)	70.34 (67.22-73.45)	62.45 (59.13-65.76)
Agri or industrial worker/daily labour/other self employed	66.56 (64.20-68.92)	67.09 (64.89-69.29)
Homemaker/housework	28.64 (27.24- 30.03)	29.84 (28.52-31.15)
Retired and unemployed (able/unable to work)	52.46 (47.72-57.19)	29.10 (24.93-33.28)
Student/Others	30.18 (27.09-33.25)	41.98 (38.63-45.34)
Wealth Index		
Richest	29.02 (26.99-31.05)	29.04 (27.16-30.92)
Richer	40.49 (38.36-42.62)	41.05 (39.06-43.05)
Middle	44.63 (42.29-46.97)	45.35 (43.15-47.54)
Poorer	50.44 (48.28-52.59)	51.51 (49.49-53.52)
Poorest	55.84 (53.59-58.09)	53.63 (51.49-55.78)

The prevalence of tobacco smoking was higher among males (61.82 %) than their female counterparts (28.67 %). Approximately same scenario was observed male adults and female adults when age was adjusted. The prevalence of smoking by educational level among adults ranged from 27.27% (Secondary school completed) to 62.29 % (no formal schooling) and when age was adjusted educational level among adults ranged from 29.26 % (Secondary school completed) to 55.92% (no formal schooling).In addition, the prevalence of smoking was highest among those in the poorest group of the wealth index (55.84 %) and lowest among those in the richest group (29.02 %). There was no difference in the group of wealth index when age was adjusted. Finally it has been found that age adjusted and unadjusted prevalence of tobacco smoking use differs only for work status which is expected (Table4.2.4)

4.3 Smokeless Tobacco Consumption

4.3.1 Current pattern of smokeless tobacco consumption (STC) in Bangladesh (GATS, 2010)

In this section, current pattern of smokeless tobacco use have been summarized in Table 4.3.1. The problem of adult smokeless tobacco consumption seemed more serious among the female (21.92% from urban areas and 30.60% from rural areas). In Bangladesh smokeless tobacco use is slightly higher among the males from rural areas (25.92%) than males from urban areas (18.13%). Of the respondents, almost equal proportions of male and female are non users.

Table 4.3.1: Prevalence of Current Smokeless Tobacco Consumption (STC)

Current smokeless tobacco users	Urban(n=4857)			Rural(n=4772)			Total(n=9629)		
	Male	Female	P-value	Male	Female	P-value	Male	Female	P-value
Daily	18.13	21.92	0.001	25.92	30.60	<0.001	21.93	26.27	<0.001
Less than daily	5.95	1.98	<0.001	6.18	1.24	<0.001	6.07	1.61	<0.001
Not at all	75.92	76.1	0.884	67.90	68.16	0.849	72.00	72.12	0.899

Note: Less than daily indicates occasionally. P-values are obtained from *Z*-test for proportion.

Figure 4.3.1 shows the patterns of STC among the adults by gender and residence.

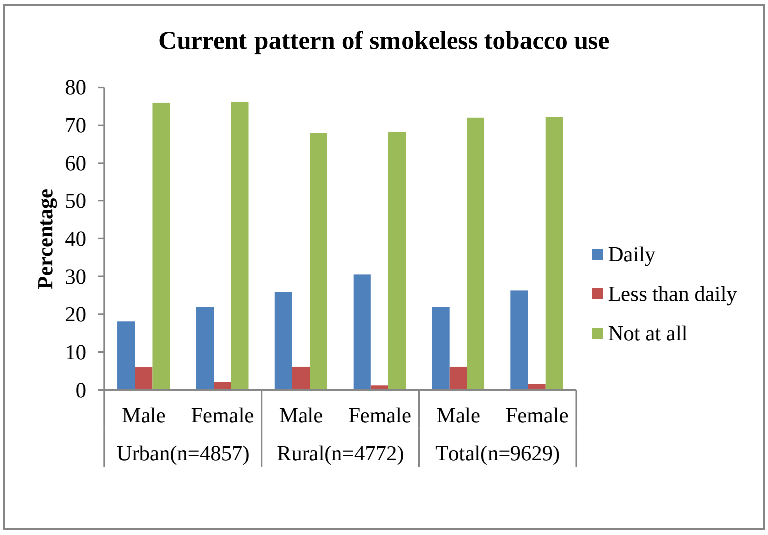

Figure 4.3.1: Current pattern of smokeless tobacco use.

4.3.2 Current smokeless tobacco consumption (STC) daily by socio-economic and demographic factors

A Comparison socio-demographic and economic characteristics of study subjects to current smokeless tobacco users has been performed (Table 4.3.2). The results showed that residence, gender, age (yrs), level of education, work status and wealth index were significantly ($p<0.001$) associated with smokeless tobacco consumption. Wealth index and level of education are gradually and significantly (p-value<0.001) decreasing the use. It is also found that age is gradually and significantly (p-value<0.001) increasing the use.

Prevalence of smokeless tobacco use is relatively higher in rural area (58.13%) than in urban area (41.87%). i.e the rural-urban STC differentials also prominent.

Table 4.3.2: Smokeless tobacco consumption daily by demographic factors (n=9629)

Socio-economic and demographic status	Current smokeless tobacco use daily		P-value*
	Yes	No	
Residence (n, %)			
Urban	978 (41.87)	3,879 (53.19)	<0.001
Rural	1,358 (58.13)	3,414 (46.81)	
Gender (n, %)			
Male	980 (41.95)	3,488 (47.83)	<0.001
Female	1,356 (58.05)	3,805 (52.17)	
Age (yrs) (n, %)			
≤ 24	96 (4.11)	1,977 (27.11)	<0.001
25-34	421 (18.02)	2,244 (30.77)	
35-45	774 (33.13)	1,763 (24.17)	
≥46	1,045 (44.73)	1,309 (17.95)	
Level of Education (n, %)			
No formal schooling	1,318 (57.08)	2,098 (28.91)	<0.001
Less than primary school completed	366 (15.85)	1,121(15.45)	
Primary school completed	254 (11.00)	861 (11.87)	
Less than secondary school completed	243 (10.52)	1,694 (23.35)	
Secondary school completed	61 (2.64)	602 (8.30)	
High school completed	29 (1.26)	434 (5.98)	
College/University completed or higher	38 (1.65)	446 (6.15)	
Work status (n, %)			
Employee (Govt. Non-Govt.)	141 (6.04)	820 (11.24)	<0.001
Business(small/large)	212 (9.08)	781 (10.71)	
Farming (land owner & farmer)	239 (10.23)	587 (8.05)	
Agri or industrial worker/daily labour/other self employed	410 (17.55)	1127 (15.45)	
homemaker/ housework	1,067 (45.68)	2,963 (40.63)	
Retired and unemployed (able/unable to work)	130 (5.57)	297 (4.07)	
Student /Other (Specify)	137 (5.86)	718 (9.85)	
Wealth Index(n, %)			
Poorest	638 (27.31)	1,228 (16.84)	<0.001
Poor	582 (24.91)	1,486 (20.38)	
Middle	424 (18.15)	1,308 (17.94)	
Rich	438 (18.75)	1,602 (21.97)	
Richest	254 (10.87)	1,669 (22.88)	

**P-values are obtained from Chi-square test (Pearson Chi-square or Likelihood Ratio whichever applicable). 95% CIs are obtained from p±1.96 se(p). – indicates N/A due to no observation. Total sample size=9629; STC=Smokeless Tobacco Consumption.

Figure 4.3.2 shows the current smokeless tobacco consumption daily by residence.

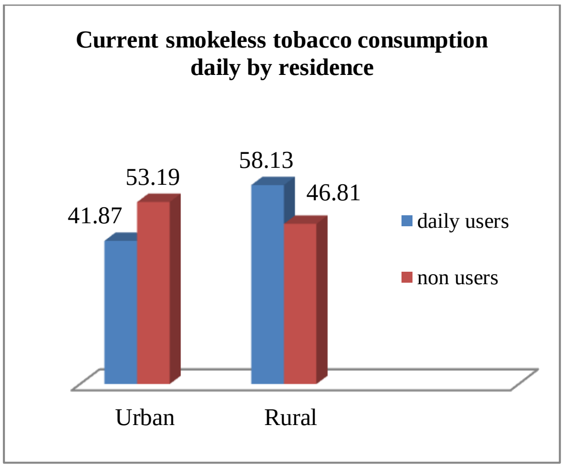

Figure 4.3.2 Current STC daily by residence

Females were more likely to consume smokeless tobacco than male (58.05% vs. 41.95%). The older age group had higher prevalence of STC than the youngsters.

Figure 4.3.3 shows the current smokeless tobacco consumption daily by gender.

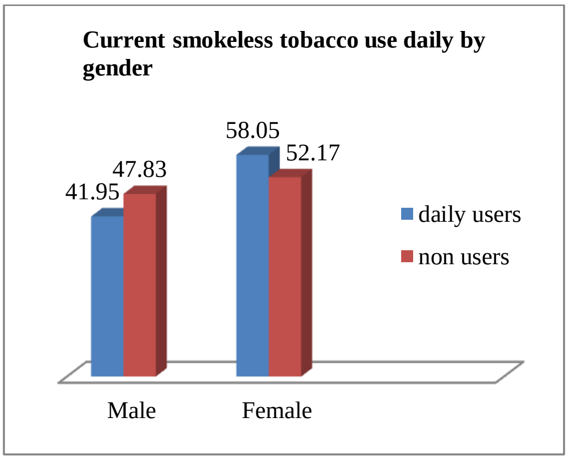

Figure 4.3.3: Current STC daily be gender

For instance, among the adults aged 45 and above, the prevalence of STC was 44.73%. In contrast, the prevalence for adults aged 15-24 years was 4.11% and adults aged 25-34 years was 18.02%. Also the prevalence for adults aged 35-44 was 33.11%. Figure 4.3.4 shows the current smokeless tobacco consumption daily by age group.

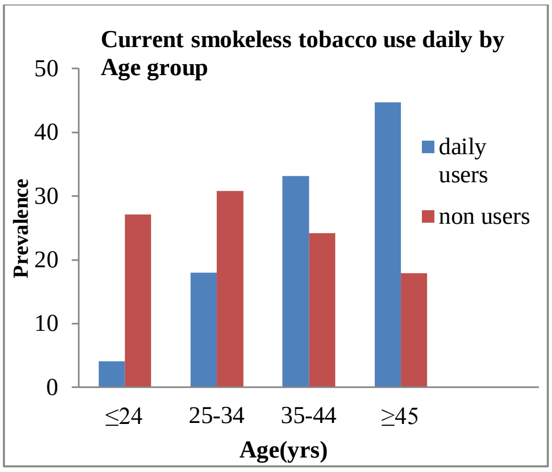

Figure 4.3.4: Current STC daily by Age group

The higher the educational level, the lower the likelihood to be smokeless tobacco users. For instance, among the adults with no formal education, 57.08% were current smokeless tobacco users. The adults who completed primary school 11% were smokeless tobacco users. Figure 4.3.5 shows the current smokeless tobacco consumption daily by educational level.

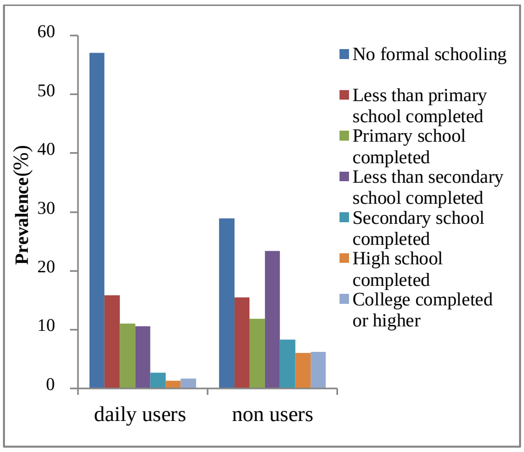

Figure 4.3.5: Current STC daily by educational level

However, the adults who completed higher education, only 1.65% were current smokeless tobacco users. Most of the smokeless tobacco users were homemaker/houseworker (45.68%). Figure 4.3.6 shows the current smokeless tobacco consumption daily by work status.

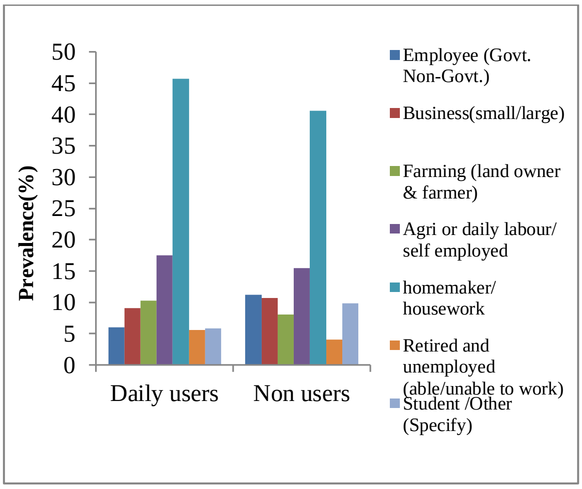

Figure 4.3.6: Current STC daily by work status

Like education wealthiest had an inverse relationship with STC. For example, among the adults in the first wealthiest (poorest) 27.31% consumed smokeless tobacco compared to 10.87% among those from the fifth quintile (richest). Also the prevalence of poor group was 24.91% and 18.75% from rich group was smokeless tobacco users. Figure 4.3.7 shows the current smokeless tobacco consumption daily by economical status (wealth index).

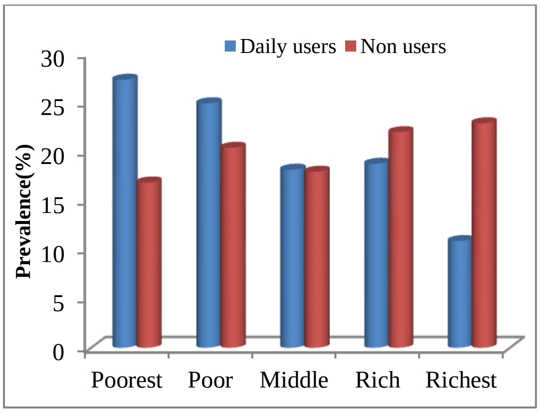

Figure 4.3.7: Current STC by economical status

4.3.3 Current pattern of smokeless tobacco product used among the users in Bangladesh stratified by area of residence and gender

The distribution of tobacco products used by residence and gender is shown in Table 4.3.3 below.

Table 4.3.3: Pattern of current smokeless tobacco use (GATS, 2010) in Bangladesh

Categories of smokeless tobacco product use		Urban (n=978)			Rural (n=1358)			Total(n=2336)		
		Male % (95% CI)	Female % (95% CI)	p-value	Male % (95% CI)	Female % (95% CI)	p-value	Male % (95% CI)	Female % (95% CI)	p-value
Snuff, by mouth (Betel quid with zarda)	daily	83.09 (79.48,86.70)	65.78 (61.86-69.69)	<0.001	76.33 (72.82-79.83)	52.53 (49.05-56.00)	<0.001	79.18 (76.64-81.73)	58.04 (55.41-60.66)	<0.001
	Less than daily	_	1.95 (0.81-3.09)	0.004	0.88 (0.11-1.65)	1.89 (0.94-2.84)	0.127	0.51 (0.06-0.96)	1.91 (1.19-2.64)	0.003
	Not at all	16.91 (13.29,20.52)	32.27 (28.41-36.13)	<0.001	22.79 (19.34-26.25)	45.58 (42.11-49.05)	<0.001	20.31 (17.79-22.82)	40.04 (37.44-42.65)	<0.001
Snuff, by nose (Betel quid with zarda)	Daily	19.32 (15.52,23.13)	33.87 (29.96-37.77)	<0.001	25.79 (22.19-29.39)	44.57 (41.11-48.03)	<0.001	23.06 (20.42-25.69)	40.42 (37.51-42.73)	<0.001
	Less than daily	1.21 (0.16,2.16)	3.19 (1.74-4.64)	0.043	0.18 (-0.17-0.52)	2.02 (1.04-3.00)	0.003	0.61 (0.12-1.10)	2.51 (1.68 -3.34)	<0.001
	Not at all	79.47 (75.58,83.36)	62.94 (58.96-66.93)	<0.001	74.03 (70.42-77.64)	53.41 (49.93-56.88)	<0.001	76.33 (73.67-78.98)	57.37 (54.74-60.01)	<0.001
Chewing tobacco (pan masala)	Daily	10.39 (7.45,13.33)	1.24 (0.33-2.15)	<0.001	7.59 (5.41-9.78)	3.41 (2.15-4.67)	0.001	8.78 (7.00-10.55)	2.51 (1.68-3.34)	<0.001
	Less than daily	1.69 (0.45,2.93)	3.90 (2.30-5.49)	0.044	-	2.90 (1.73-4.07)	<0.001	0.71 (0.19-1.24)	3.32 (2.37-4.37)	<0.001
	Not at all	87.92 (84.78-91.06)	94.86 (93.04-96.68)	<0.001	92.40 (90.22-94.59)	93.69 (91.99-95.38)	0.356	90.51 (88.68-92.35)	94.17 (92.93-95.42)	0.001

Table 4.3.3 Cont.

Categories of smokeless tobacco product use		Urban (n=978)			Rural (n=1358)			Total(n=2336)		
		Male % (95% CI)	**Female % (95% CI)**	**p-value**	**Male % (95% CI)**	**Female % (95% CI)**	**p-value**	**Male % (95% CI)**	**Female % (95% CI)**	**p-value**
Sadapata chewing	Daily	7.97 (5.36-10.58)	1.42 (0.44-2.34)	<0.001	7.42 (5.26-9.58)	3.16 (1.94-4.37)	<0.001	7.65 (5.99-9.32)	2.43 (1.61-3.25)	<0.001
	Less than daily	1.93 (0.61-3.26)	3.55 (2.02-5.07)	0.135	-	3.16 (1.94-4.37)	<0.001	0.82 (0.25-1.38)	3.32 (2.37-4.27)	<0.001
	Not at all	90.10 (87.22-92.97)	95.04 (93.24-96.83)	0.003	92.58 (90.42-94.74)	93.69 (91.99-95.38)	0.423	91.53 (89.79-93.27)	94.25 (93.01-95.49)	0.011
Gul	Daily	27.78 (23.46-32.09)	16.49 (13.43-19.55)	<0.001	20.85 (17.50-24.19)	16.41 (13.83-18.99)	0.037	23.78 (21.11-26.44)	16.45 (14.47-18.42)	<0.001
	Less than daily	0.97 (0.02-1.91)	3.19 (1.74-4.64)	<0.001	-	3.41 (2.15-4.67)	<0.001	0.41 (0.00-0.80)	3.32 (2.37-4.27)	<0.001
	Not at all	71.25 (66.89-75.61)	80.32 (77.04-83.60)	0.001	79.15 (75.81-82.49)	80.18 (77.40-82.95)	0.643	75.82 (73.14-78.49)	80.24 (78.12-82.36)	0.010
Khoinee	Daily	9.42 (6.61-12.23)	1.77 (0.68-2.86)	<0.001	6.36 (4.35-8.37)	1.77 (0.85-2.69)	<0.001	7.65 (5.99-9.32)	1.77 (1.07-2.47)	<0.001
	Less than daily	0.97 (0.02-1.91)	3.01 (1.60-4.43)	0.029	-	3.03 (1.84-4.22)	<0.001	0.41 (0.01-0.81)	3.02 (2.11-3.94)	<0.001
	Not at all	89.61 (86.67-92.55)	95.21 (93.45,96.97)	0.001	93.64 (91.63-95.65)	95.20 (93.71-96.69)	0.211	91.94 (90.23-93.64)	95.21 (94.07-96.34)	0.001
Any	Daily	7.00 (4.55-9.46)	2.30 (1.07-3.54)	<0.001	5.83 (3.90-7.76)	1.52 (0.66-2.37)	<0.001	6.33 (4.80-7.85)	1.84 (1.13-2.56)	<0.001
	Less than daily	0.73 (-0.09-1.54)	3.72 (2.16-5.29)	0.003	0	3.03 (1.84-4.22)	<0.001	0.31 (-0.04-0.65)	3.32 (2.37-4.27)	<0.001
	Not at all	92.27 (89.69-94.84)	93.97 (92.00-95.94)	0.295	94.17 (92.24-96.10)	95.45 (94.00-96.91)	0.288	93.37 (91.81-94.93)	94.84 (93.66-96.02)	0.133

Any products include: Betel quid with supari, Betel quid with zarda, Pan Masala, Sada pata, Sada pata with dry Banana leaf mixture, Sada pata with neam leaf mixture, Tobacco powder, use Tobacco leaf for brushing teeth, Zarda, Zarda with Supari. '-' no observations

Female respondents of 27.02% used smokeless tobacco only and male respondents of 13.54% used smokeless tobacco only. It was found that most used smokeless tobacco product was betel quid with zarda (79.18% male and 58.04% female). There was a difference in the pattern of using smokeless tobacco for males versus females – 23.06% male used betel quid with zarda (snuff by nose) compared to paan masala (8.78%), sada pata chewing (7.65), Gul (23.78%), khoinee (7.65%) or other products (6.33%) whereas female users (40.42%) used betel quid wth zarda (snuff by nose), paan masala (2.51%), sadapata chewing (2.43%), Gul (16.45%), khoinee (1.77%) or other products include (1.84%). The distribution of tobacco products by male and female are shown in the Figure 4.3.8 below.

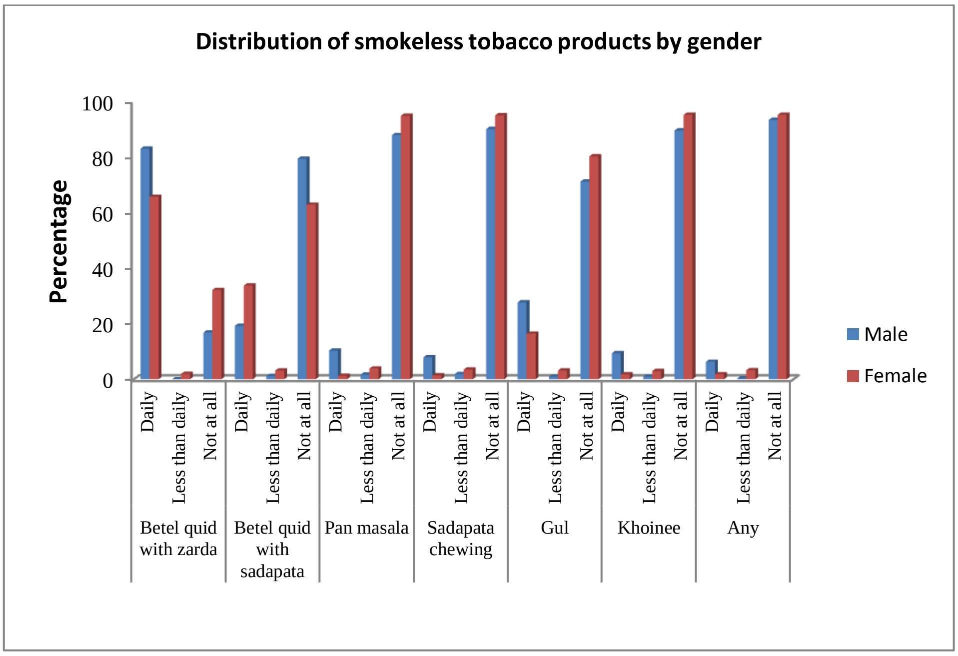

Figure 4.3.8: Distribution of smokeless tobacco consumption by male and female

83.9% male from urban area used betel quid with zarda whereas 65.78% female from rural area used betel quid with zarda (snuff by mouth). Male and female respondents from urban area used betel quid with zarda most (83.09% vs. 65.78%). 76.33% male whereas 52.53% female from rural area used betel quid with zarda. 19.32% male from urban area whereas 25.79% male from rural area used snuff by nose (betel quid with zarda).

33.87% female from urban area whereas 44.57% female from rural area used snuff by nose (betel quid with zarda).

The distribution of tobacco products by male and female are shown in the Figure 4.3.9 below.

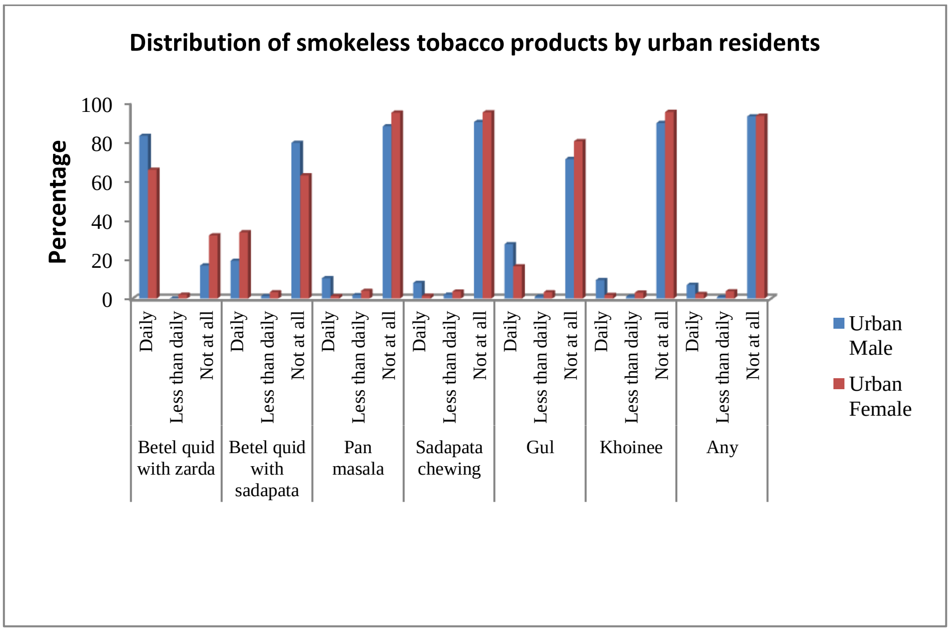

Figure 4.3.9: Pattern of tobacco consumption by urban residents

Paan masala was used daily by 10.39% male from urban area whereas 7.59% male from rural area and 1.24% female from urban area whereas 3.41% female from rural area.

The distribution of tobacco products by rural respondents are shown in the Figure 4.3.10 below.

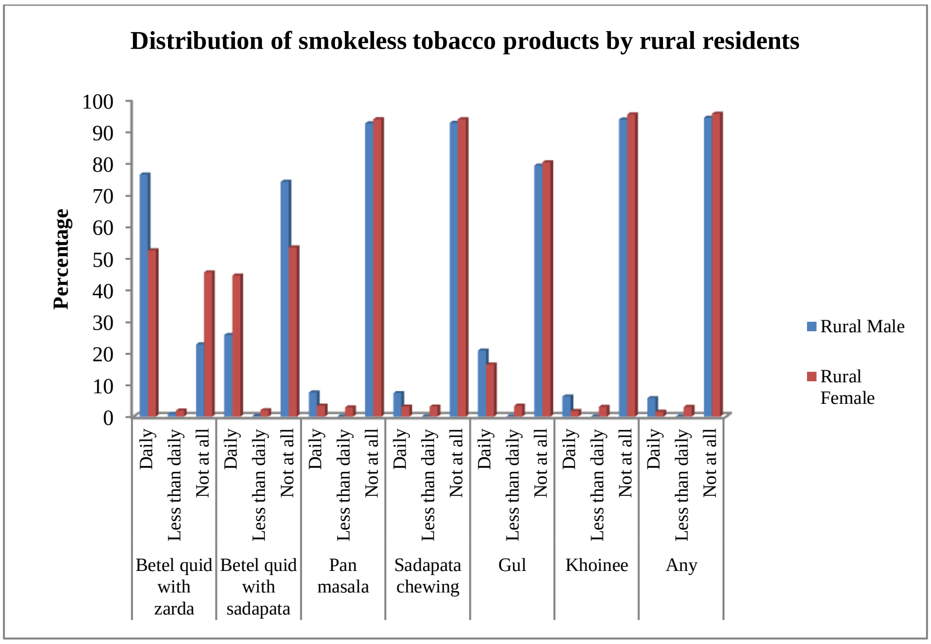

Figure 4.3.10: Pattern of tobacco consumption by rural respondents

Gul was used daily by 27.78% male from urban area whereas 20.85% male from rural area and 16.49% female from urban area whereas 16.41% female from rural area. Khoinee was another smokeless tobacco product used daily by 9.42% male from urban area whereas 6.36% male from rural area and 1.77% female used from both area. Also other types of smokeless tobacco products were consumed by 6.33% of male and 1.84% of female. Of the male tobacco users, over 33.82% were smokers, 14.46% were mixed users, and only 13.54% were smokeless users, 38.18% were non users; however, female tobacco users showed the opposite pattern –27.20% were smokeless users, 0.68% were mixed users, and 0.79% were smokers, 71.32% non users.

4.3.4 Age adjusted and unadjusted prevalence of smokeless tobacco use

Age adjusted and unadjusted prevalence of STC among adults by selected socio-demographic characteristics is presented in Table 4.3.4. It has been found that age adjusted and unadjusted prevalence of smokeless tobacco use differs only for work status which is expected. Prevalence of STC among the urban adults was 20.14% and for rural adults was 28.46%. Approximately same result was observed when age was adjusted.

Table 4.3.4: Age adjusted and unadjusted prevalence of STC

Socio-economic and demographic status	Prevalence (95%CI) Unadjusted	Prevalence (95% CI) Age Adjusted
Residence		
Urban	20.14 (19.01- 21.26)	21.04 (19.94-22.14)
Rural	28.46 (27.18-29.74)	27.42 (26.23-28.60)
Gender		
Male	21.93 (20.72-23.15)	21.04 (19.94-22.14)
Female	26.27 (25.07-27.47)	27.42 (26.23-28.59)
Educational Level		
College completed or higher	7.85 (5.45-10.25)	8.00 (5.66-10.35)
High school completed	6.26 (4.06-8.07)	7.79 (5.18-10.40)
Secondary school completed	9.20 (7.00-11.40)	11.83 (9.19- 14.46)
Less than secondary school completed	12.55 (11.07- 14.02)	16.01 (14.27- 17.76)
Primary school completed	22.78 (20.32-25.24)	25.89 (23.33-28.44)
Less than primary school completed	24.61 (22.42-26.80)	25.24 (23.12-27.36)
No formal schooling	38.58 (36.95-40.21)	32.60 (31.10-34.09)
Work Status		
Employee (Govt. Non-Govt.)	14.67 (12.44-16.91)	16.96 (14.69-19.24)
Business (small/large)	21.35 (18.80 -23.90)	20.90 (18.63-23.17)
Farming (land owner and farmer)	28.93 (25.84- 32.03)	22.32 (19.88-24.75)
Agri or industrial worker/daily labour/other self employed	26.68 (24.46-28.89)	28.07 (25.99-30.15)
Homemaker/housework	26.48 (25.11-27.84)	27.69 (26.40-28.98)
Retired and unemployed (able/unable to work)	30.44 (26.08-34.81)	12.03 (9.72--4.33)
Student/Others	16.03 (13.56-18.48)	24.51 (21.3- 27.68)
Wealth Index		
Richest	13.21 (11.70-14.72)	13.36 (11.92- 14.81)
Richer	21.47 (19.69-23.25)	21.89 (20.19-23,59)
Middle	24.48 (22.46-26.51)	25.06 (23.13-26.98)
Poorer	28.14 (26.20-30.08)	29.12 (27.28-30.97)
Poorest	34.19 (32.04-36.34)	31.93(29.96-33.90)

*STC=Smokeless Tobacco Consumption

The prevalence of tobacco smoking was higher among males (21.93%) than their female counterparts (26.27%). Approximately same scenario was observed for male adults and female adults when age was adjusted. The prevalence of smoking by educational level among adults ranged from 7.85% (completed college or higher) to 38.58% (no formal schooling) and when age was adjusted educational level among adults ranged from 8% (completed college or higher) to 32.60% (no formal schooling). In addition, the prevalence of smoking was highest among those in the poorest group of the wealth index (28.14%) and lowest among those in the richest group (13.21%). There was no difference in the group of wealth index when age was adjusted. Finally it has been found that age adjusted and unadjusted prevalence of smokeless tobacco use differs only for work status which is expected (Table 4.3.4).

4.4 Discussion

The study revealed that the adults aged 15 years or above living in rural areas had higher tendency of using tobacco. Also in accordance with prevalences, almost all male adults were tobacco smokers compared to female smokers. Older people consumed tobacco most. With increasing level of education the use of tobacco is decreasing. Age is gradually and significantly increasing the use of tobacco. Poor adults have a significant inverse relation with tobacco use Pattern of using tobacco products was different from male to female. It was also found that the male adults in urban areas used Manufactured cigarettes most. Betel quid with zarda was most consumed smokeless tobacco products among females. Other products like gul, bidis, zarda are comparatively less consumed.

CHAPTER FIVE

BINARY LOGISTIC REGRESSON TO DETERMINATE TOBACCO USE

CHAPTER FIVE

BINARY LOGISTIC REGRESSON TO DETERMINATE TOBACCO USE

5.1 Introduction

In the previous chapter the comparison of the characteristics subjects to tobacco consumption has been performed and the corresponding graphs have been presented. Also, distribution of tobacco products using z-test for proportion and chi-square test have been performed. In the present chapter analyze the socio-demographic and economic determinants to current tobacco by binary logistic regression. Obtained results by logistic regression analysis has been interpreted using odds ratio (OR) and 95% confidence interval (CI). Adjusted and unadjusted Odds Ratio has been reported with OR and 95% confidence interval, too.

5.2 Analyze the socio-demographic and economic predictors to current tobacco smoking using binary logistic regression

Adjusted and Unadjusted relationships of socio-demographic and economic variables have been analyzed to find the predictors of current tobacco smoking (Table 5.2.1) using binary logistic regression. The binary logistic regression was estimated where the response variable takes the value of "1" if the respondent is a daily or less than daily tobacco smokers and "0" otherwise.

The results from the binary logistic regression analysis in Table 5.2.1 showed that respondents of rural area are significantly and more likely to tobacco smoking (unadjusted OR=1.16, 95% CI=1.05- 1.27; adjusted OR=0.90, 95% CI=0.78-1.03). Males are significantly and more likely to tobacco smoking currently (unadjusted OR=62.45, 95% CI=49.42-78.91; adjusted OR=36, 95% CI=25.30-52.82). Older group (greater than or equal 45 years) (unadjusted OR=3.28, 95% CI=2.78-3.86) are 3 times more likely to tobacco smoking than youngest group (less than or equal to 24 years).Approximately same scenario was observed when age was adjusted. Respondents of age groups 25-34 years are 2 times more likely (unadjusted OR=2.24, 95% CI=1.90-2.64; adjusted OR=2.18, 95% CI=1.79-2.67) and respondents of age group 35-45 years are 3 times more likely (unadjusted OR=3.18, 95% CI=2.71-3.74) and 2 times more likely (adjusted OR=2.46, 95% CI=2.01-3.01) to smoking tobacco than youngest age group. Respondents with no formal schooling are most likely to smoked tobacco than respondents of other educational level for both adjusted and unadjusted model. Work status does not seem to have any significant pattern of influence to tobacco smoking for adjusted model, although in unadjusted model respondents of some work status are found to be more likely to use tobacco smoking than government and non-government employee. The poorest group (Unadjusted OR=2.09, CI= 1.79-2.45 and Adjusted OR=1.80, CI=1.39-2.33) and poorer group (Unadjusted OR=2.02, CI= 1.73-2.37 and Adjusted OR=1.50, CI=1.18-1.91) were 2 times more likely to be tobacco

Table 5.2.1: Socio-economic and demographic correlates to current tobacco smoking

Socio-economic and demographic status	Unadjusted OR(95%CI)	Adjusted OR(95%CI)
Residence		
Urban (RC)	1.00	1.00
Rural	1.16(1.05-1.27)	0.90(0.78-1.03)
Gender		
Female (RC)	1.00	1.00
Male	62.45(49.42-78.91)	36(25.30-52.82)
Age group(years)		
≤ 24 (RC)	1.00	1.00
25-34	2.24(1.90-2.64)	2.18(1.79-2.67)
35-45	3.18(2.71-3.74)	2.46(2.01-3.01)
≥ 46	3.28(2.78-3.86)	2.53(2.05-3.12)
P-value for trend	<0.001	<0.001
Educational Level		
College completed or higher	1.00	1.00
High school completed	0.69(0.48-0.98)	1.05(0.72-1.55)
Secondary school completed	0.79(0.58-1.09)	1.36(0.95-1.93)
Less than secondary school Completed	0.99(0.76-1.28)	1.99(1.47-2.72)
Primary school completed	0.96(0.73-1.27)	2.05(1.46-2.89)
Less than primary school completed	1.63(1.26-2.11)	2.64(1.90-3.66)
No formal schooling	1.91(1.50-2.44)	3.81(2.77-5.26)
P-value for trend	<0.001	<0.001
Work Status		
Employee (Govt. Non-Govt.)	1.00	1.00
Business (small/large)	2.49(2.06-3.01)	1.28(1.03-1.58)
Farming (land owner and farmer)	3.04(2.49-3.71)	1.01(0.80-1.29)
Agri or industrial worker/daily labour/ other self employed	2.63(2.21-3.13)	1.22(0.97-1.53)
Homemaker/housework	0.03(0.02-0.04)	0.39(0.24-0.62)
Retired and unemployed (able/unable	0.82(0.63-1.07)	0.57(0.42-0.78)
Student/Others	0.54(0.43-0.68)	0.82(0.63-1.09)
P-value for trend	<0.001	<0.001
Wealth Index		
Richest (RC)	1.00	1.00
Rich	1.44(1.23-1.70)	1.10(0.89-1.36)
Middle	1.62(1.37-1.90)	1.24(0.89-1.56)
Poor	2.02(1.73-2.37)	1.50(1.18-1.91)
Poorest	2.09(1.79-2.45)	1.80(1.39-2.33)
P-value for trend	<0.001	<0.001
p-Value from Hosmer - Lemeshow goodness of ft test=0.169, AUC=0.891.		

Note: RC=Reference Category; OR=Odds Ratio; CI=Confidence Interval; P-values for trend are for more than three categories; AUC, Area Under ROC Curves is for prediction accuracy of the adjusted model; Hosmer- Lemeshow test and AUC is for goodness of ft of the adjusted model.

smokers than richest group. Respondents who belongs to richer group (Unadjusted OR=1.44, CI=1.23-1.70 and Adjusted OR=1.10, CI=0.89-1.36) were most likely to tobacco smoking than respondents of richest group.

5.3 Analyze the socio-demographic and economic predictors to current smokeless tobacco using binary logistic regression

Adjusted and Unadjusted relationships of socio-demographic and economic variables have been analyzed to find the predictors of current smokeless tobacco use (Table 5.3.1) using binary logistic regression. The binary logistic regression was estimated where the response variable takes the value of "1" if the respondent is a daily smokeless tobacco user and "0" otherwise.

The results from the binary logistic regression analysis in Table 5.3.1 showed that respondents of rural area are significantly and more likely to use smokeless tobacco (unadjusted OR=1.58, 95% CI=1.44-1.73; adjusted OR=1.13, 95% CI=1.00-1.26). Females are significantly and more likely to use smokeless tobacco currently (unadjusted OR=1.27, 95% CI=1.15-1.39; adjusted OR=1.70, 95% CI=1.41-2.05).

Table 5.3.1: Adjusted and unadjusted odds ratios derived from logistic regression to current smokeless tobacco use (daily) in Bangladesh

Socio-economic and demographic factors	Unadjusted OR (95% CI)	Adjusted OR (95% CI)
Residence		
Urban (RC)	1.00	1.00
Rural	1.58 (1.44-1.73)	1.13 (1.00-1.26)
Gender		
Male (RC)	1.00	1.00
Female	1.27 (1.15-1.39)	1.70 (1.41-2.05)
Age group (years)		
≤24 (RC)	1.00	1.00
25-34	3.86 (3.07-4.86)	3.61 (2.85-4.57)
35-44	9.04 (7.24-11.28)	8.48 (6.72-10.71)
≥45	16.44 (13.19-20.49)	14.72 (11.61-18.64)
P-value for trend	<0.001	<0.001
Educational Level		
College completed or higher	1.00	1.00
High school completed	0.78 (0.48-1.29)	1.04 (0.62-1.75)
Secondary school completed	1.19 (0.78-1.82)	1.61 (1.03-2.50)
Less than secondary school completed	1.68 (1.18-2.41)	2.23 (1.51-3.28)
Primary school completed	3.46 (2.42-4.96)	3.75 (2.52-5.59)
Less than primary school completed	3.83 (2.69-5.45)	3.38 (2.28-5.02)
No formal schooling	7.37 (5.26-10.34)	4.37 (2.96-6.44)
P-value for trend	<0.001	<0.001
Work Status		
Employee (Govt. Non-Govt.)	1.00	1.00
Business (small/large)	1.58 (1.24-1.99)	0.97 (0.75-1.27)
Farming (land owner and farmer)	2.37 (1.87-2.99)	0.84 (0.64-1.11)
Agri or industrial worker/daily labour/other self employed	2.12 (1.71-2.62)	0.88 (0.69-1.13)
Homemaker/housework	2.09 (1.73-2.54)	0.72 (0 .56-0.94)
Retired and unemployed (able/unable to work)	2.55 (1.94-3.34)	0.83 (0.60-1.14)
Student/Others	1.11 (0.86-1.43)	1.12 (0.83-1.51)
P-value for trend	0.001	<0.001
Wealth Index		
Richest (RC)	1.00	1.00
Richer	1.79 (1.52-2.13)	1.32 (1.09-1.61)
Middle	2.13 (1.79-2.53)	1.39 (1.12-1.70)
Poorer	2.57 (2.19-3.03)	1.64 (1.33-2.02)
Poorest	3.41 (2.90-4.02)	1.79 (1.44-2.21)
P-value for trend	<0.001	<0.001
p-Value from Hosmer - Lemeshow goodness of ft test=0.223, AUC=0.773.		

Note: RC=Reference Category; OR=Odds Ratio; CI=Confidence Interval; P-values for trend are for more than three categories; AUC, Area Under ROC Curves is for prediction accuracy of the adjusted model; Hosmer- Lemeshow test and AUC is for goodness of ft of the adjusted model.

Older group (greater than or equal 45 years) are 16 times more likely to use smokeless tobacco than youngest group (less than or equal to 24 years) (unadjusted OR=16.44, 95% CI=13.1- 20.49). For adjusted model, the older group is found to be about 15 times more likely to use than the youngest group

(OR=14.72, 95% CI=11.61- 18.61). Respondents of age groups 25-34 years are 3 times more likely (unadjusted OR=3.86, 95% CI=3.07- 4.86; adjusted OR=3.61, 95% CI=2.85- 4.57) and respondents of age group 35-44 years are 8 times more likely (unadjusted OR=9.04, 95% CI=7.24-11.28 and adjusted OR=8.48, 95% CI=6.72- 10.71) to use smokeless tobacco than youngest age group. Respondents with no formal schooling are most likely to use smokeless tobacco than respondents of all educational level for both adjusted and unadjusted model. Work status does not seem to have any significant pattern of influence to smokeless tobacco use for adjusted model, although in unadjusted model respondents of some work status are found to be more likely to use smokeless tobacco than government and non-government employee. The poorest group (Unadjusted OR=3.41, CI= 2.90-4.02 and Adjusted OR=1.79, CI=1.44-2.21) and poorer group (Unadjusted OR=2.57, CI= 2.19-3.03 and Adjusted OR=1.64, CI=1.33-2.02) were 2 times more likely to be smokeless tobacco users than richest group. Respondents who belongs to richer group (Unadjusted OR=1.79, CI=1.52-2.13 and Adjusted OR=1.32, CI=1.09-1.61) were most likely to use smokeless tobacco than respondents of richest group.

5.4 Discussion

The study revealed that the adults aged 15 years or above living in rural areas had higher tendency of using tobacco. Male adults were more likely to be tobacco users overall, and were more likely to smoke than women, while female adults were more likely to be smokeless users or non users than males. Both smoked and smokeless tobacco consumption revealed a strong inverse relation with age and level of education. Tobacco consumptions were found to be higher among older adults with no formal education compared to those with higher education. Poor people were more likely to use tobacco (smoked or smokeless) than rich people.

CHAPTER SIX

MULTILEVEL MODEL TO DETERMINATE TOBACCO USE IN BANGLADESH

CHAPTER SIX

MULTILEVEL MODEL TO DETERMINATE TOBACCO USE IN BANGLADESH

6.1 Introduction

In the previous chapter binary logistic regression analysis was performed to identify the predictors of tobacco smoking and smokeless tobacco using. Many demographic, socio-economic and tobacco related factors are significantly associated with tobacco consumption. Current pattern of tobacco consumption was performed in the previous chapter. In the previous chapter age adjusted and unadjusted prevalence and Odds Ratio and 95% CI from Binary Logistic Regressions for Factors Influencing tobacco consumption among Adults was performed. Logistic regression is very useful for situation in which we want to be able to predict the presence or absence of a characteristic or out come to bare on values of a set of predictor variable. It is similar to a linear regression but is suited to mode where the variable is dichotomous and the independent variable may be either dummy or categorical. We have used binary logistic regression model to identify the significant variables, which have important effects to current tobacco consumption. In this chapter, Multilevel modeling has been developed in response to the challenge of appropriately analyzing clustered data. Because of the complex structure of the model and the nature of the error terms, multilevel models are estimated using iterative Empirical Bayes/maximum likelihood (EB/ML) techniques, rather than the OLS methods typically employed to estimate the parameters of single-level models. The standard error estimates for a multilevel model are more accurate than those for a single-level individual-as-unit-of-analysis model. In addition to the correction of standard error estimates and the more appropriate significance tests that result, multilevel models also provide other advantages over traditional analytic techniques. Prominent among these is the ability to simultaneously examine the effects of variables at both individual and group levels, as well as possible cross-level interaction effects (Krull JL & MacKinnon 2001; Bryk & Raudenbush, 1992).

6.2 Multilevel Modeling

The GATS, Bangladesh-2010 data set used in this study was based on multistage cluster sampling. For this reason, the hierarchical structure of the data creates the dependence among observations. Hence, observations within a same cluster are correlated. With a view to taking into consideration the clustering effect in the data, we considered multilevel modeling (Hardin and Hilbe, 2012), which accounts the correlation among the observations within a cluster.

6.3 Multilevel modeling to tobacco smoking

In this section multilevel (two- level) logistic regression has been performed. Where the response variable was current tobacco smoking and the covariates are sex, age (yrs), educational level, work status and wealth index. Measures of association (odds ratio) and measures of variance (intra-class correlation (ICC)) were calculated, as well as the discriminatory accuracy by calculating the area under the ROC curve (AUC). The first step examined the null model of overall probability of tobacco smoking without adjustment for predictors.

6.3.1 Association of various cofactors with tobacco smoking

Association of various cofactors with tobacco smoking has been analyzed using chi-square test (Pearson or likelihood ratio whichever applicable).

Table 6.3.1: Test of association: chi-square tests of independence between explanatory and dependent variable

Explanatory Variable	χ^2	Explanatory Variable	χ^2
Sex	2.9e+03***	Wealth Index	110.49***
Age (yrs)	259.67***		
Level of education	202.07***		
Work status	2.5e+03***		

Note: ***p<0.0001

Apparently all the factors seem to be very much influential on the dependent variable, current tobacco smoking. Table 6.3.1 represents the results of test of association between current tobacco smoking and each of the explanatory variables. All explanatory variables are found highly statistically associated with dependent variable.

6.3.2 Intercept Only Logistic Model

Table 6.3.2: Odds ratio and standard errors of an intercept-only logit model and intercept-only multilevel models predicting the probability of tobacco smoking

Model effect	Single level logistic regression		Multilevel logistic regression	
Fixed effect	**S.E.**	**Odds ratio**	**S.E.**	**Odds ratio**
Intercept	0.007	0.302	0.01	0.297
Random effect	-			
Intercept (level-2), var (S.E.)			0.04 (0.02)**	
ICC(%)			1.63	
-2logL (Deviance)	10429.48		10420.22	

**p-value<0.001.

The ML estimate from the single level logistic model of the ratio of current tobacco smokers to current tobacco nonuser is 0.30. It is in fact odds-ratio when no predictors have been considered in the model. In comparison, the same ratio is estimated to be 0.297, from the multilevel model by the adaptive Gauss-Hermite quadrature methods respectively. Compared to the odds ratios obtained by all multilevel methods the standard logistic model odds-ratio has been overestimated. It is observed that there is a significant difference between the standard logistic estimate and the multilevel logistic estimate. Table 6.3.2 shows that the random effects i.e. the cluster specific effects are significant at 5% level of significance. Therefore, failing to take into account the standard logistic model has overestimated the odds-ratio by about 1.68% [(0.302-0.297)*100/0.297] when multilevel model by corresponding methods adaptive Gauss-Hermite quadrature has been applied (see Table 6.3.2).

6.3.3 Multilevel Logistic Regression Model

The model is followed with all the significant factors to assess their simultaneous affect on tobacco smoking. The ICC is 0.063 (Table 6.3.3), which indicates 6 percent of the total variance in tobacco smoking is explained at the cluster-level. It is observed that there exist significant differences between the odds ratios of these two models for each of the explanatory variables.

Table 6.3.3 presents odds ratios and 95% CI from a single level logistic model predicting the probability of tobacco smoking and its equivalent from multilevel model. The last two columns of Table 6.3.3 represent respectively the difference in odds ratio between single and multilevel multivariate models and percentage of under or overestimation of odds ratio by single level modeling. Male respondents were 44.17 times more likely to smoked tobacco product than females. The multilevel model shows that the probability of tobacco smoking increases significantly with age, adjusting for the effect of other predictors and respondents of age greater than 45 are 2.63 times more likely to smoked tobacco product than youngest group (age <25 years),whereas under the single level model the corresponding odds is 2.53 times higher. Thus the odds ratio has also underestimated significantly by about 3.80%. The multilevel effect is observed notably for predictor gender. The odds ratio under single level model is highly overestimated (18.49%) compared to multilevel estimates. Level of education seems to be another influential factor in regulating tobacco smoking except for the highest level of education (high school completed).Among those respondent who have no formal schooling and less than primary education the respective odds of tobacco smoking is about 3.93 times and 2.74 times higher compared to the odds of tobacco smoking among respondents without any formal education for the multilevel model, whereas under the single level model the corresponding odds ratio are 3.81 times and 2.64times higher, respectively.

It is also found that work status is not a significant predictor of tobacco smoking. Results reveal that wealth index (WI) or respondent's economic status is another significant correlates of tobacco smoking. The probability of tobacco smoking is low among the respondents who are from economically well off families. The multilevel analysis shows that the respondents from middle, poor and poorest economic status have OR of tobacco smoking 1.79%, 1.55% and 1.23% higher compared to the odds among richest group of respondents. The corresponding figures under single level model are about 1.80%, 1.50% and 1.24%. Thus for wealth index the odds ratios of

Table 6.3.3: Identifying correlates of tobacco smoking in Bangladesh using multilevel logistic regression analysis

Socio-economic and demographic correlates	Single level logistic regression	Multilevel logistic regression	Odds ratio Difference	Over/Under Estimation (%)
	OR (95% CI)	OR (95% CI)		
Gender				
Female (RC)	1.00	1.00		
Male	36(25.30-52.82)	44.17 (30.13- 64.74)	-8.17	18.49
Age (yrs)				
≤ 24 (RC)	1.00	1.00	-	-
25-34	2.18(1.79-2.67)	2.21 (1.79-2.73)	-0.03	1.36
35-45	2.46(2.01-3.01)	2.54 (2.06-3.13)	-0.08	3.15
≥46	2.53(2.05-3.12)	2.63 (2.11-3.27)	-0.1	3.80
Level of Education				
College/University completed or higher (RC)	1.00	1.00	-	-
High school completed	1.05(0.72-1.55)	1.02 (0.68-1.52)	0.03	2.94
Secondary school completed	1.36(0.95-1.93)	1.37 (0.95-1.98)	-0.01	0.73
Less than secondary school completed	1.99(1.47-2.72)	2.04 (1.48-2.82)	-0.05	2.45
Primary school completed	2.05(1.46-2.89)	2.08 (1.46-2.99)	-0.03	1.44
Less than primary school completed	2.64(1.90-3.66)	2.74 (1.94-3.87)	-0.1	3.65
No formal schooling	3.81(2.77-5.26)	3.93 (2.80-5.50)	-0.12	3.05
Work status				
Employee (Govt. Non-Govt.) (RC)	1.00	1.00	-	-
Business(small/large)	1.28(1.03-1.58)	1.30 (1.04-1.63)	-0.02	1.54
Farming (land owner & farmer)	1.01(0.80-1.29)	1.01 (0.78-1.79)	0.00	0
Agri or industrial worker/daily labour/other self employed	1.22(0.97-1.53)	1.24 (0.98-1.57)	-0.02	1.61
homemaker/ housework	0.39(0.24-0.62)	0.42 (0.26-0.69)	-0.03	7.14
Retired and unemployed (able/unable to work)	0.57(0.42-0.78)	0.56 (0.40-0.77)	0.01	1.79
Student /Other (Specify)	0.82(0.63-1.09)	0.81(0.61-1.09)	0.01	1.23
Wealth Index				
Richest (RC)	1.00	1.00	-	-
Rich	1.10(0.89-1.36)	1.09 (0.88-1.36)	0.01	0.92
Middle	1.24(0.89-1.56)	1.23 (0.96-1.56)	0.01	0.81
Poor	1.50(1.18-1.91)	1.55 (1.21-1.98)	-0.05	3.23
Poorest	1.80(1.39-2.33)	1.79 (1.37-2.34)	0.01	0.56

Table 6.3.3 Cont.

	Single level logistic regression	Multilevel logistic regression	Odds ratio Difference	Over/Under Estimation (%)
Cluster variance		0.223** (SE=0.05)		
MOR		1.57		
No. of observation		9565		
No. of cluster		399		
Intraclass Correlation (ICC, %)		6.34	-	-
AUC	0.8911	0.9076		
Log-likelihood	-3181.27	-3161.33		
AIC	6406.55	6366.66		
BIC	6564.20	6524.31		
Daviance	6362.55	6322.66		

Note:*p<0.05; **p<0.001; CI=Confidence Interval; OR=Odds Ratio; RC=Reference Category, Intercept is not shown in the table. MOR=Median Odds Ratio,

middle, poor and poorest group have been overestimated respectively by 0.81%, 3.23% and 0.56%.Thus it is evident that if multilevel effect is not taken into account in the model, the estimates would be either underestimated or overestimated considerably. These result simply that single-level model for this data set is not appropriate. Also multilevel model performs better than single level model.

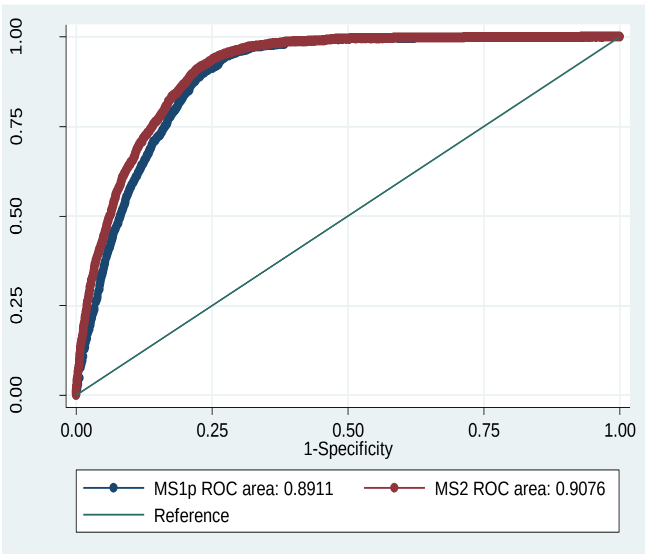

Figure 6.1(a): Area under the receiver operating characteristics (AUC) curve for tobacco smoking plotted separately for single level logistic regression model (Blue thick line) and for multilevel model (Red thick line) which adjusted for sex, age, education, occupation, wealth index. The diagonal line represents an AUC equal to 0.50.

The ROC analysis demonstrated a high level of predictive accuracy, with an area under the ROC curve (AUC) of 0.8911 for multilevel model whereas AUC of the ROC curve is 0.9076 for single level model (Figure 6.1(a), Table 6.3.3). Comparing both AIC and BIC statistics (Table 6.3.3) it is clear that the multilevel logistic regression model is preferable to the simple logistic regression model.

6.4 Multilevel modeling to smokeless tobacco use

Multilevel analysis produces more valid results when lower levels (individuals) are nested within higher levels (Gatscluster i.e. 400 mouza/mohallah). Measures of association (odds ratio) and measures of variance (intra-class correlation (ICC)) were calculated, as well as the discriminatory accuracy by calculating the area under the Receiver Operator Characteristic (AUROC) curve. Instead of standard single level logistic model, multilevel logistic regression model has been utilized since the data follow a hierarchical structure (Figure 2.1). Also the comparison between single and multilevel model has been done to investigate the necessity of multilevel effects. The findings suggest that sex, age, education, religion, and wealth index have significant multilevel effects on smokeless tobacco use.

6.4.1 Association of various cofactors with smokeless tobacco consumption

Association of various cofactors with tobacco smoking has been analyzed using chi-square test (Pearson or likelihood ratio whichever applicable).

Table 6.4.1: Test of association: chi-square tests of independence between explanatory and dependent variable

Explanatory Variable	χ^2	Explanatory Variable	χ^2
Sex	24.5513***	Wealth Index	253.6286***
Age (yrs)	1.1e+03***		
Level of education	764.3602***		
Work status	118.5910***		

Note: ***p<0.0001, **p<0.008

Apparently all the factors seem to be very much influential on the dependent variable, currently smokeless tobacco use daily. Table 6.4.1 represents the results of test of association between currently smokeless tobacco use daily and each of the explanatory variables. All explanatory variables are found highly statistically associated with dependent variable.

6.4.2 Intercept Only Logistic Model

An intercept-only model that predicts the probability of smokeless tobacco use is fitted first. Table 6.4.2 represents the odds ratio and standard error of single level and multilevel logistic regression model.

Table 6.4.2. Odds ratio and standard errors of an intercept-only logit model and intercept-only multilevel models predicting the probability of smokeless tobacco use

Model effect	Single level logistic regression		Multilevel logistic regression	
Fixed effect	**S.E.**	**Odds ratio**	**S.E.**	**Odds ratio**
Intercept	0.008	0.320	0.01	0.289
Random effect	-			
Intercept (level-2), var (S.E.)			0.42 (0.05)	
ICC(%)			11.42	
-2logL (Deviance)	10670.07		10393.94	

6.4.3 Multilevel Logistic Regression Model

The multilevel logistic model is followed with all the significant factors to assess their simultaneous affect on smokeless tobacco use. The ICC is .10, which indicates 10 percent of the total variance in using smokeless tobacco is explained at the cluster-level. It is observed that there exist significant differences between the odds ratios of these two models for each of the explanatory variables. Table 6.4.3 presents odds ratios and 95% CI from a single level logistic model predicting the probability of smokeless tobacco use and its equivalent multilevel model. The last two columns of this table represent respectively the difference in odds ratio between single level and multilevel multivariate model and percentage of under

Table 6.4.3: Identifying correlates of smokeless tobacco use in Bangladesh using multilevel logistic regression analysis

Socio-economic and demographic correlates	Single level logistic regression OR (95% CI)	Multilevel logistic regression OR (95% CI)	Odds ratio difference	Over/Under Estimation (%)
Gender				
Male (RC)	1.00	1.00		
Female	1.70 (1.41-2.05)	1.72** (1.39, 2.07)	-0.02	1.17
Age (yrs)				
≤ 24 (RC)	1.00	1.00	-	-
25-34	3.61** (2.85-4.57)	3.73 ** (2.93, 4.75)	-0.12	3.32
35-45	8.48** (6.72-10.71)	9.09**(7.16, 11.54)	-0.61	7.19
≥46	14.72** (11.61- 18.64)	16.09**(12.60, 20.53)	-1.37	9.30
Level of Education				
College/University completed or higher (RC)	1.00	1.00		
High school completed	1.04 (0.62-1.75)	1.08 (0.63-1.83)	-0.04	3.85
Secondary school completed	1.61* (1.03-2.50)	1.71* (1.08- 2.71)	-0.1	6.21
Less than secondary school completed	2.23** (1.51-3.28)	2.32** (1.55- 3.49)	-0.09	4.04
Primary school completed	3.75** (2.52-5.59)	3.91** (2.57- 5.95)	-0.16	4.27
Less than primary school completed	3.38** (2.28-5.02)	3.61** (2.38- 5.46)	-0.23	6.80
No formal schooling	4.37** (2.96-6.44)	4.93** (3.28- 7.41)	-0.56	12.81
Work status				
Employee (Govt. Non-Govt.) (RC)	1.00	1.00		
Business(small/large)	0.97 (0.75-1.27)	0.92 (0.70-1.21)	0.05	5.15
Farming (land owner & farmer)	0.84 (0.64-1.11)	0.83 (0.63-1.12)	0.01	1.19
Agri or industrial worker/daily labour/other self employed	0.88 (0.69-1.13)	0.82 (0.64-1.09)	0.06	6.82
homemaker/ housework	0.72 (0 .56-0.94)	0.71 (0.54-0.94)	0.01	1.39
Retired and unemployed (able/unable to work)	0.83 (0.60-1.14)	0.77 (0.55-1.08)	0.06	7.23
Student /Other (Specify)	1.12 (0.83-1.51)	0.95 (0.69-1.31)	0.17	15.18
Wealth Index				
Richest (RC)	1.00	1.00		
Rich	1.36* (1.09- 1.61)	1.32* (1.07-1.63)	0.04	2.94
Middle	1.39** (1.12-1.70)	1.33**(1.07-1.66)	0.06	4.31
Poor	1.64**(1.33-2.02)	1.54** (1.23- 1.92)	0.10	6.09
Poorest	1.79** (1.44-2.21)	1.67** (1.33- 2.09)	0.12	6.70

Table 6.4.3 Cont.

	Single level logistic regression	Multilevel logistic regression	Odds ratio difference	Over/Under Estimation (%)
Cluster variance		0.374** (SE=0.05)		
MOR		1.79		
No. of observation		9565		
No. of cluster		399		
Intraclass Correlation (ICC, %)		10.2	-	-
AUC	0.773	0.8187		
Log-likelihood	-4433.17	-4349.72		
AIC	8907.13	8743.44		
BIC	9043.28	8901.09		
Daviance	8869.13	8699.44		

Note:*p<0.05; **p<0.001; CI=Confidence Interval; OR=Odds Ratio; RC=Reference Category, Intercept is not shown in the table.

or overestimation of odds ratio by single level multilevel modeling. Male respondents were 1.72 times more likely to be smokeless tobacco users than females. The multilevel model shows that the probability of using smokeless tobacco increases significantly with age, adjusting for the effect of other predictors. The odds ratio under single level model is slightly overestimated (1.17%) compared to multilevel estimates. That is, the multilevel effect is observed slightly for predictor gender. Level of Education seems to be another influential factor in regulating smokeless tobacco use except for the highest level of education (high school completed). Among those respondent who have no formal schooling and less than primary education the respective odds of using smokeless tobacco is about 3.61 times and 4.93 times higher compared to the odds of smokeless tobacco use among respondents without education for the multilevel model whereas under the single level model the corresponding odds is 3.38 times and 4.37 times higher respectively. The odds of smokeless tobacco use among respondents of age group $\geq$46 is about 16.09 times higher than the odds among respondents of age group $\leq$24 under the multilevel model whereas under the single level model the corresponding odds is 14.72 times higher. Thus the odds ratio has also underestimated significantly by about 9.30%. We also showed that work status was not significant predictors of smokeless tobacco use. Table 6.4.3 reveals that wealth index (WI) or respondent's economic status is another significant determinant of smokeless tobacco use. The probability of smokeless tobacco use is low among the respondents who are from economically well off families. The multilevel analysis shows that the respondents from middle, poor and poorest economic status have odds of smokeless tobacco use 33%, 54% and 67% higher compared to the odds among richest respondents. The corresponding figures under single level model are about 39%, 64% and 79%. Thus for wealth index the odds ratios of category middle, poor and poorest have been overestimated respectively by 4.31%, 6.09% and 6.70%.Thus it is evident that if multilevel effect is not taken into account in multilevel modeling the estimates would be either underestimated or overestimated considerably. These results imply that single-level model for this outcome variable is not appropriate. Also multilevel model performs better than single level model. The ROC curve demonstrated a high level of predictive accuracy, with an area under the curve of 0.8187 (Fig 6.1(b)).

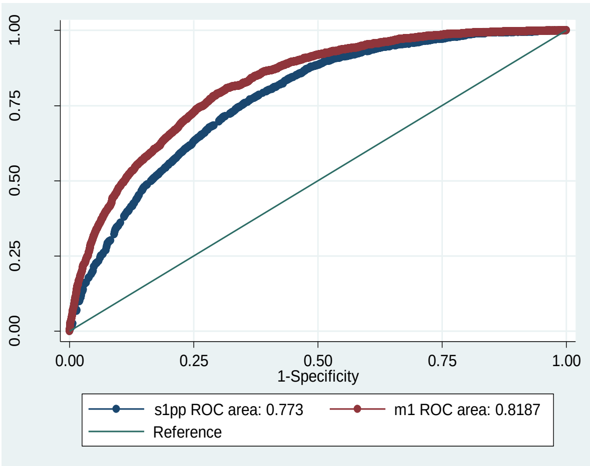

Figure 6.2(b): Area under the receiver operating characteristics (AUC) curve for the use of smokeless tobacco plotted for separately for single level logistic regression model (Blue thick line) and for multilevel model (Red thick line) which adjusted for sex, age, education, occupation, wealth index. The diagonal line represents an AUC equal to 0.50.

for multilevel model whereas AUC of the ROC curve is 0.773 (Fig 6.2(b)) for single level model (simple logistic regression). Comparing both AIC and BIC statistics in Table 6.4.3 it is clear that the multilevel logistic regression model is preferable to the simple logistic regression model.

6.5 Discussion

We have presented multilevel logistic regression analysis to estimate individuals and cluster influences on tobacco smoking and smokeless tobacco using. We analyse two different individual outcomes (currently tobacco smoke or not and currently use smokeless tobacco or not) for which the relative importance of cluster influences differs substantially (Merlo J et. al., 2016). This chapter revealed that current tobacco smoking and smokeless tobacco use was strongly influenced by socio-economic and demographic characteristics (sex, age, educational level, wealth index). Our study provides concepts and innovative analytical approaches like the use of the AUC that allow improved multilevel analysis of tobacco use.

CHAPTER SEVEN

DISCUSSION AND CONCLUSION

CHAPTER SEVEN

DISCUSSION AND CONCLUSION

7.1 Discussion

Tobacco use is considered to be a venerable problem in Bangladesh. This study is based on a country representative cross-sectional study of Bangladesh. The study findings from both single level and multilevel analyses demonstrate that the prevalence of smoking and smokeless tobacco varied significantly by gender, age, education levels, wealth index, and occupational categories. About 65.11% of male and 34.89% of female were tobacco users. Overall prevalence of adults tobacco consumption in Bangladesh was male. Respondents who were in the age group 45 years or older were 16.44 times more likely to use smokeless tobacco and twice more likely to smoked tobacco compared to those who were 15-24 years. Tobacco consumption was greatly influenced by the level of education. Adults with no education led to 1.91 times and 7.37 times higher likelihood to tobacco smoking and smokeless tobacco use respectively. Majority of the adults (54.29%) resided in rural areas. A positive association was found between tobacco consumption (smoked and smokeless tobacco products) and wealth index. The respondents from poorest families were 1.80 times more likely to smoke tobacco compared to those who were richest. Likewise, the odds ratio was about 1.79 times for those who were use smokeless tobacco. Various tobacco products, whether smoked, smokeless were popular among the adults in Bangladesh. About 66.18% of the male respondents used manufactured cigarettes. The prevalence rates of *bidi* smoking were about 42.79% for male and 75.76% for female. The prevalence of using other tobacco products betel quid with zarda was 79.18% for male adults. The prevalence rates of betel quid with zarda (snuff, by nose) and gul were roughly the same (about 23%) by male. This analysis indicated that, in

Bangladesh, the use of tobacco is more common among older adults aged 46+ years, living in rural areas, with lower socioeconomic status, and less education for both genders. The present study found no statistically significant difference in the prevalence of smoking between the urban and rural populations, but did find that the use of smokeless tobacco was statistically higher in the rural population. In general tobacco use was increasing with increasing age. As expected, men were more likely to use tobacco than female. Men were much more likely to smoke (because smoking would be very socially unacceptable by female) and smokeless tobacco use was more likely to be seen in women (not as undesirable as smoking). Increasing educational level and wealth status had an overall defensive effect for all form of tobacco use with a few exceptions. In addition, adults with low level of education, and from poor households had significantly higher risk of tobacco consumption. Prevalence of smoking manufactured cigarettes was more among male and urban residents. Although smoking is not so prevalent in female groups, among the smoked tobacco products bidi was smoked most by female respondents. Our study also revealed that occupational status was not Prevalence of smokeless tobacco use of this study reveals that one in every five men and one in every four women currently use smokeless tobacco daily/occasionally in Bangladesh. The use was more prevalent in rural area and among older subjects. Subjects with no formal schooling and of lowest wealth index were most likely to use smokeless tobacco. However, some exceptions were seen. In a few countries increased wealth and education were not associated with decreased tobacco use, with Mexico actually having lower tobacco use in the poor, and with the lowest rates of tobacco use in China present in the poorest and wealthiest (Palipudi KM et al, 2015). For wealth index, odds ratios were computed taking the highest wealth category as reference. The trend (decreasing odds of tobacco use with increasing wealth) was significant for Bangladesh and also for India (Singh A et al., 2014). From our study we find that manufactured cigarettes, bidis (smoking) and betel quid with zarda, gul (smokeless) are most consumed tobacco products. The fact that over one-fourth of men and women in

Bangladesh used tobacco in some form or other is also a source of concern. For women smoking cigarettes or *bidis* is considered socially unacceptable in the South-East Asian community, but using smokeless tobacco is socially acceptable (Rani M et al., 2018).The observation that smokeless tobacco use increased with age is consistent with previous reports. Also rural peoples in neighboring country like India (Rani M et al., 2003) used chewing substances more as is found in this study for Bangladesh. Our finding shows that male smoked tobacco most and female used smokeless tobacco most which is identical to other study (Palipudi, et al., 2012). Similar pattern had been found for Bangladeshi community in UK (Giovino GA et al., 2012). Compared to developed countries, the tobacco consumption issue in the South East Asian Region (SEAR) is much more complex. This is because of several reasons such as wide range of products with varying components, unregulated and unorganized market forces, and production and sale in informal settings. Our study provides information about prevalence and patterns of smokeless tobacco use among men and women in Bangladesh not covered in other multicounty surveys and confirm that tobacco use was higher among women, the less educated, and the poor, particularly those living in rural areas. Prevalence of consuming betel quid with zarda (snuffs by mouth) was more among urban and male population. This concluding chapter provides a brief summary of the findings, policy recommendations, and the strength and limitations of the study. In addition, it also explains the contributions of the study and provides suggestions for future research. The findings of this research have implications for tobacco control policies, cessation strategies and interventions — to be more effective, these strategies need to account for smokeless tobacco, gender, and social norms. Compared with smoking, smokeless tobacco consumption and prevention has been a more neglected policy area. As noted by a recent gathering of experts on smokeless tobacco policies in South Asia (Gupta PC et al., 2003), any existing policies on smokeless tobacco are either inadequate or poorly enforced, and there is a need for greater coordination of policies and improvement of existing legislation (Khan et al., 2014). Though the current study describes the pattern of tobacco smoking and smokeless tobacco use among various socioeconomic and demographic sub-groups, attention should be given to further investigating reasons for tobacco use in Bangladesh to develop an evidence base for interventions.

7.2 Policy Implication

Although Bangladesh has tobacco control laws and policies, tobacco consumption among the adults is common. The study on adults tobacco consumption have identified several factors significantly associated to tobacco consumption among the adults and suggested some guidelines for policy purpose.

- A greater proportion of the tobacco users in Bangladesh were male. Therefore, comprehensive strategies along with preventive programs should be tightened to help male smokers avoid smoking, As expected, male tobacco users in Bangladesh were more likely than female tobacco users to be smokers, whereas there were more female smokeless users than male smokers overall. In Bangladesh in particular, there is a need to take a more gendered perspective in tobacco control, which has largely been missing thus far in most of the world (Amos et al., 2012).
- Because of the higher use of smokeless tobacco among women in Bangladesh, tobacco cessation strategies for smokeless tobacco must also address the specific needs of women. As older people from rural areas smoked tobacco most, tobacco cessation policies should be increased in the region.
- Adults from poor families in Bangladesh are more likely to be vulnerable, comprehensive control strategies should be implemented to these groups to overcome these problems. Manufactured cigarettes and bidis were most consumed smoked tobacco products urban and rural people respectively. Also Betel quid with zarda and gul were most

consumed smoked tobacco products. Considering the above issues in mind, to reduce the epidemic of diseases caused by smoked and smokeless tobacco use, we need to take action to reduce the use of these products. Policymakers need to consider smokeless tobacco use separately in tobacco control efforts, since the economic and health effects of smokeless tobacco use are different from that of smoking. Given the wide acceptance of smokeless tobacco use in Bangladeshi culture, interventions to raise awareness of the harms of smokeless tobacco use should target women, especially older as a key group.

- Poorer smokers with low education smoke more. Tobacco companies deliberately market brands toward those people. Tobacco control interventions could be made more effective among these groups. At the same time, national campaign programs should focus on changing social norms by addressing inappropriate attitudes and perceptions of risk towards smokeless tobacco use among the Bangladeshis.
- Similar to the effective policies for tobacco smoking, policies such as increasing excise tax and restricting marketing of smokeless tobacco to targeted population including minors, should also be initiated. These interventions should complement the existing intervention strategies aimed at reducing smokeless tobacco use among the public in Bangladesh.
- An important finding in this study is the high prevalence rate of tobacco consumption among the older age group (45 and above) that demands more attention. Therefore, targeting termination of tobacco consumption in these age groups would be extremely important for policy formulation.
- The finding suggests that improvement of education could be an important strategy for reducing both tobacco use. Therefore, increasing knowledge through education on the harmful effects of tobacco and changing attitude towards tobacco consumption through counseling programs could be good interventions. Religious leaders especially

Imams (the head of a mosque) could play a vital role in preventing tobacco consumption.

7.3 Contribution of the Study

Prevalence, patterns, and determinants of tobacco consumption among adults are commonly researched in this study. Several studies also reported on tobacco consumption, but such studies are very much limited for product specifications. Many of the studies dealt with small data sets, which may not be representative. Due to lack of national-level data in many developing countries, little is known about the vulnerability of these issues. Therefore, this study contributes in several ways to the literature, methodological approach, and policy recommendations. This thesis has important contributions in literature. The use of nationally representative data enhances the literature on Bangladesh about prevalence, patterns and determinants of tobacco consumption. Further, Pattern of using different types of tobacco products (both smoked and smokeless) is considered in this study. This thesis has expanded the applications of statistical techniques, thus contributing to the methodological approach. Different options of dependent variables were also included in the analysis. Most of the studies on tobacco consumption employed binary logistic regressions (BLRs) for analysis. Multilevel logistic regressions were used to find out the determinants of tobacco consumption. It is important to monitor smoking rates as well as smokeless tobacco use in Bangladesh to prevent more women from switching to smoked tobacco or initiating smoking at a young age. Reduce most consumed smoked tobacco products rates in male and smokeless tobacco products rates in female would also have implications for the present research. Since detailed gender and product specific analysis were conducted in this study, policymakers should be benefited. Longitudinal surveys and cohort studies are recommended for examining these policy related issues.

7.4 Strengths and Limitations of the Study

GATS was the first large-scale survey ever conducted in Bangladesh that used electronic devices such as handheld computers, often called the Pocket PC systems. This device was useful to facilitate the complex skip pattern used in the GATS questionnaire, as well as some in-built validity checks on questions during data collection. A repeated quality control mechanism was used to test the quality of questionnaire programming. The Bangladesh Demographic and Health Survey (BDHS) was conducted by the National Institute of Population Research and Training (NIPORT) of MoHFW with the consultation of Mitra and Associates, a Bangladeshi research firm and Macro International Inc. The standard ethical clearance was obtained from the country and informed consent from respondents was taken during the survey. For reliability, the household head or any senior knowledgeable person of the household was selected to collect information. Therefore GATS produced representative and independent cross-sectional information for each country. In addition, the use of different statistical tools such as graphs, bivariate analysis, multilevel analysis and comparisons of the results produced by different techniques are also the strengths of this study. Multilevel analysis using hierarchical data have not been carried out in Bangladesh on tobacco using yet. The theoretical, conceptual and analytical frameworks and the existing literature guided the selection of dependent and independent variables for the study. However, there are several limitations that need to be addressed. The findings in this report are based on self-reports. Furthermore, education categories were combined into broad groupings, which could have contributed to biased estimates in terms of the gradients observed. Nonetheless, these groupings provided greater precision than those used in earlier tobacco use research in Bangladesh. The data used in constructing wealth index is based on limited number of asset variables and was not constructed separately for urban and rural, which might result in incomplete or under representing socioeconomic status. All tobacco product use is self-reported and may be subject to recall bias. Study design allowed for the investigation of only a limited number of socio-demographic variables.

Some other variables like psychological variables could provide more predicting accuracy, but no such variable was available. Finally, since the datasets are cross-sectional in nature, cause-effect relationships could not be inferred. Despite these limitations, the current study provides evidence of the significance of social determinants on tobacco use. Findings indicate that social determinants and their role should be given high precedence when addressing the issue of tobacco use.

7.5 Conclusion

Tobacco use is one of the foremost causes of preventable morbidity and mortality. Our study reveals that tobacco smoking is more prevalent among male and smokeless tobacco use is prevalent among female. Smoking rates is remain lower in women than in men and prevalence is particularly low in women in Asia but is high in Poland, Russia and the UK. Besides Bangladesh smokeless tobacco use is particularly prevalent in India. Our study also reveals that the prevalence of tobacco use is generally higher among rural, less educated and low economic groups. We also found that the prevalence of use increased as age increased for all forms of tobacco use. An important finding in this study is high prevalence of tobacco use in the middle ages (45 years and above). The health effects of tobacco use start becoming apparent in these age groups in a major way. Therefore, targeting cessation in these age groups would be extremely important as a component of overall policy initiatives for reducing tobacco use prevalence. This will be crucial in reducing morbidity and mortality caused by tobacco use in the immediate future. In general, social determinants such as education and wealth were correlated with increased tobacco use.

In conclusion, the findings provide evidence that socio-economic and demographic factors were associated with tobacco consumption behaviours among adults. Giving it as public health priority, WHO Framework Convention on Tobacco Control should be implemented. In addition, a nationwide campaign is needed to educate people in rural area about the health risks of tobacco use.

BIBLIOGRAPHY

Bibliography

Abdullah AS, Hitchman SC, Driezen P, Nargis N, Quah ACK & Fong GT (2011). Socioeconomic differences in exposure to tobacco smoke pollution (TSP) in Bangladeshi households with children: Findings from the International Tobacco Control (ITC) Bangladesh survey. *International Journal of Environmental Research and Public Health*, *8*(3), 842-860.

Achia TNO (2015) Tobacco Use and Mass Media Utilization in Sub-Saharan Africa. *PLOS ONE*, Vol. 10, Issue 2: e0117219. doi:10.1371/journal.pone.0117219.

Acquah H (2010). Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship. *Journal of Development and Agricultural Economics,* Vol. 2(1) pp. 001-006.

Afulani PA (2015) Rural/Urban and Socioeconomic Differentials in Quality of Antenatal Care in Ghana. *PLoS ONE* 10(2): e0117996. doi:10.1371/journal.pone.0117996

Agresti, A. (2007). An introduction to categorical data analysis. Hoboken, NJ: John Wiley & Sons

Amos A, Hiscock R, Fidler JA, Bauld L and Munafo M (2012). Socioeconomic status and smoking: a review. *ANNALS OF THE NEW YORK ACADEMY OF SCIENCES*, ISSN 0077-8923. Available at: http://www.who.int/tobacco/surveillance/survey/gats/en/. Access date: December 28, 2016.

Aziz S, Choudhury T, Huque N (2015). Tobacco Use in Rural Bangladesh: Socio-Demographic Characteristics, Self Rated Health and Diseases. *Public Health Research*, Vol.5, pp 198-205.

Bender R & Grouven U (1997). Ordinal logistic regression in medical research. *Journal of Royal College of Physicians*, *31*(5), 546-551.

Bhise MD, Patra S (2018) Prevalence and correlates of hypertension in Maharashtra, India: A multilevel analysis. PLoS ONE 13(2): e0191948. https://doi.org/10.1371/journal.pone.0191948

Bickel P, Zhang P (1992). Variable selection in nonparametric regression with categorical covariates. *J. Am. Stat. Assoc*. 87: 90–97.

Bozdogan H (1987). Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extensions. Psychometrika (52)3: 345-370.

Bush J, White M, Kai J, et al. (2003). Understanding influences on smoking in Bangladeshi and Pakistani adults: community based, qualitative study. *BMJ Primary Care*; 326:962.

Campbella P, Jordana KP, Smith BH, Scotland G , Dunna KM (2018). Chronic pain in families: a cross-sectional study of shared social, behavioural, and environmental influences. January 2018·Vol. 159, Number 1, pp.41–47.

Chan, Y. H. (2004). Biostatistics 202: Logistic regression analysis. *Singapore Medical Journal*, *45*(4), 149-153.

Chen Y H, Yeh CY, Chen RY, Chien LC, Yu PT, Chao KY, Han BC(2009). Moving toward people's needs for smoke-free restaurants: Before and after a national promotion program in Taiwan, 2003–2005. *Nicotine & Tobacco Research*, *11*(5), 503-513.

Corsi DJ, Lear SA, Chow CK, Subramanian SV, Boyle MH, Teo kk (2013). Socioeconomic and Geographic Patterning of Smoking Behavior in Canada: A Cross-Sectional Multilevel Analysis. *PLOS ONE*, Vol.8, Issue 2: e57646. doi:10.1371/journal.pone.0057646.

Cox DR and Hinkley DV (1974). Theoretical Statistics. CHAPMAN & HALL, New York, Washington DC.

D. Efroymson, S. Ahmed, J. Townsend, S. M. Alam, A. R. Dey, R. Saha, et al. (2001). "Hungry for tobacco: an analysis of the economic impact of tobacco consumption on the poor in Bangladesh," *Tobacco control*, vol. 10, pp. 212-217.

Dhungana RR, Khanal MK, Baniya A (2013). Prevalence of Current Tobacco Use Among Secondary Students in Nepal: A Meta-Analysis. *Journal of Chitwan Medical College*; Vol. 3 Issue 6: pp 1-5.

Do YK and Bautista MA (2015). Tobacco use and household expenditures on food, education, and healthcare in low- and middle- income countries: a multilevel analysis. *BMC Public Health*, Vol-15, pp 1098. http://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-015-2423-92/15. Access on: March 12, 2017.

Elton-Marshall T, Fong GT, Zanna MP, Jiang Y, Hammond D, O'Connor, et al. (2010). Beliefs about the relative harm of "light" and "low tar" cigarettes: Findings from the International Tobacco Control (ITC) China Survey. *Tobacco Control*; Vol. 19 Suppl 2: i54-62

Evans-Whipp TJ, Bond L, Ukoumunne OC, Toumbourou JW and Catalano RF (2010). The Impact of School Tobacco Policies on Student Smoking in Washington State, United States and Victoria, Australia. *International Journal of Environmental Research and Public Health*, Vol.7, pp 698-710.

Ezzati M, Lopez AD (2003). Estimates of global mortality attributable to smoking in 2000. *Lancet*.; Vol. 362(9387), pp:847–852.

Flora MS, Mascie-Taylor CGN, Rahman M (2009). Gender and locality differences in tobacco prevalence among adult Bangladeshis. *Tobacco Control*; Vol. 18, pp:445–450. doi:10.1136/tc.2008.028142

Gaete J, Ortúzar C, Zitko P, Montgomery A and Araya R (2016). Influence of school-related factors on smoking among Chilean adolescents: a cross-sectional multilevel study. *Biomedical Central of Pediatric*. Vol-16, pp 76. doi: 10.1186/s12887-016-0612-z

Gilani SI, Leon DA (2013). Prevalence and sociodemographic determinants of tobacco use among adults in Pakistan: findings of a nationwide survey conducted in 2012. *Population Health Metrics*, 11:16. http://www.populationhealthmatrics.com/content/11/1/16 . Access date: July 13, 2015.

Goldstein H, Browne W, Rasbash J (2002). Partitioning variation in generalised linear multilevel models. *Understanding Statistics*; Vol. 1, pp:223–32.

Guo Q, Johnson CA, Unger JB, Lee L, Xie B, Chou CP, Palmer PH, Sun P, Gallaher P, Pentz MA (2007). Utility of the theory of reasoned action and theory of planned behavior for predicting Chinese adolescent smoking. *Addictive Behaviors*, Vol.32, pp 1066–1081.

Gupta PC, Ray CS (2003). Smokeless tobacco and health in India and South Asia. *Respirology*; Vol. 8, pp 419–431.

Hakim S, Chowdhury MAB, Uddin MJ (2017) Correlates of unsuccessful smoking cessation among adults in Bangladesh. *Preventive Medicine Reports* 8, pp. 122–128.

Hardin JW & Hilbe JW (2012). "Generalized Linear Models and Extensions, 3rd Edition," *Stata Press books*, *StataCorp LP*, edition 3, number glmext, April.

Hauck W and Donner A (2012). Wald's Test as applied to Hypothesis n Logit Analysis. *Journal of the American Statistical Association*, Vol. 72, pp 851-853.

Hausman J & McFadden D (1994). Specification tests for the multinomial logit model. *Econometrica, 52*(5), 1219-1240.

Heo J, Oh J, Subramanian SV, Kawachi I (2014). Household and School-Level Influences on Smoking Behavior among Korean Adolescents: A Multilevel Analysis. *PLoS ONE* 9(6): e98683. doi:10.1371/journal.pone.0098683

Hosmer DW and Lemeshow S (1989) Applied Logistic Regression. John Wiley & Sons, Inc., New York.

Hosmer, D.W. and Lemeshow, S. (2000) Applied logistic regression. 2nd Edition, John Wiley & Sons, Inc., New York. doi:10.1002/0471722146

Hossain MS, Kypri K, Rahman B, Arsian I, Akter S, Milton AH (2014). Prevalence and correlates of Smokeless Tobacco Consumption among Married Women in Rural Bangladesh. PLOS ONE, Vol. 9, Issue 1: e84470.doi:10.1371/journal.pone.0084470.

Hossain S, Hossain S, Ahmed F, Islam R, Sikder T, Rahman A (2017). Prevalence of Tobacco Smoking and Factors Associated with the Initiation of Smoking among University Students in Dhaka, Bangladesh. *Central Asian Journal of Global Health*, Volume 6, No. 1, ISSN 2166-7403.

Hu T-W, Mao Z, Liu Y, Beyer JD, Ong M (2005). Smoking, standard of living, and poverty in China. *Tobacco Control*, Vol. 14, pp 247-250.

Huang H-L, Chen F-L, Hsu C-C, Yen Y-Y, Ted Chen, Huang C-M, Shi H-Y, Hu C-Y and Lee C-H (2010). A multilevel-based study of school policy for tobacco control in relation to cigarette smoking among children in elementary schools: gender differences. *Health Education Research,* Vol.25, no.3, pp 451–463.

Huisman, M., Kunst, A., & Mackenbach, J. (2005). Inequalities in the prevalence of smoking in the European Union: Comparing education and income. *Preventive Medicine*, *40*, 756-764.

Huque R, Zaman MM, Huq SM, Sinha DN (2017). Smokeless tobacco and public health in Bangladesh. *Indian J Public Health*; Vol.61,pp:S18-24.

Imtiaz D, Kandpal SD, Juyal R (2015). A cross sectional study on quitting behavior of tobacco use among rural population in Dehradun, Uttarakhand. Indian Journal of Community Health; Vol.27, Issue no 1: pp 30-34.

ITC Bangladesh Summary, Promoting Evidence-Based Strategies to Fight the Global Tobacco Epidemic, 2010.

Jafarabadi MA, Allahverdipour H, Bashirian S, and Jannati A (2012). Modeling the Underlying Predicting Factors of Tobacco Smoking among Adolescents. *Iranian Journal of Public Health*, Vol. 41, Issue 5: pp 46–57.

Jha P, Peto R, Zatonski W, Boreham J, Jarvis MJ, et al. (2006) Social inequalities
in male mortality, and in male mortality from smoking: indirect estimation from national death rates in England and Wales, Poland, and North America. *Lancet* 2006; 368: 367–70.

Kass RE, Raftery A (1995). Bayes Factors. *J. Am. Stat. Assoc*. 90: 773-795.

Khan MH, Khan A, Kraemer A and Mori M (2009). Prevalence and correlates of smoking among urban adult men in Bangladesh: slum versus non-slum comparison. *BMC Public Health*, Vol. 9: page -149, doi:10.1186/1471-2458-9-149.

Khan MHR, Shaw JE (2011). Determinants of Contraceptive Use in Bangladesh: A Hierarchical Modeling Approach. *The International Journal of Interdisciplinary Social Sciences*: *Annual Review*, Vol.4, Issue 4· DOI: 10.18848/1833-1882/CGP/v04i04/52877.

Khayyati F, Taymoori P, Mohammadpoorasl A, Allahverdipour H, Asghari Jafarabadi M (2016). Underlying Predictors of Tobacco Smoking among Iranian Teenagers: Generalized Structural Equation Modeling. *International Journal of Pediatric*; Vol-4, pp 3461-73. DOI: 10.22038/ijp.2016.7383.

Kishore J, Jena PK, Bandyopadhyay C, Swain M, Das S, Banerjee I (2013). Hardcore Smoking in Three South-East Asian Countries: Results from the Global Adult Tobacco Survey. *Asian Pacific Journal of Cancer Prevention*, Vol. 14, Issue 2, pp 625-630.

Kristjansson AL, Sigfusdottir ID, and Allegrante JP (2013). Adolescent substance use and peer use: a multilevel analysis of cross-sectional population data. *Substance Abuse Treatment, Prevention and Policy*. Vol-8, pp 27. https://substanceabusepolicy.biomedcentral.com/articles/10.1186/1747-597X-8-27. Access date: November 01, 2016.

Krull JL & MacKinnon DP (2001). Multilevel Modeling of Individual and Group Level Mediated Effects, *Multivariate Behavioral Research*, 36:2, 249-277, DOI: 10.1207/S15327906MBR3602_06

Kuha J (2004). AIC and BIC: Comparisons of Assumptions and Performance. Sociol. Methods Res. (33)2: 188-229.

Larsen K, Petersen JH, Budtz-Jorgensen E, Endahl L. Interpreting parameters in the logistic regression model with random effects. Biometrics. 2000; 56(3):909–14. PMID: 10985236

Leatherdale ST and Manske S (2005). The Relationship Between Student Smoking in the School Environment and Smoking Onset in Elementary School Students. *Cancer Epidemiology, Biomarkers and Prevention*, Vol.14, pp 1762-1765.

Manimunda SP, Benegal V, Sugunan AP, Jeemon P, Balakrishna N, Thennarusu K, Pandian D and Pesala KS (2012). Tobacco use and nicotine dependency in a cross-sectional representative sample of 18,018 individuals in Andaman and Nicobar Island, India. *BMC Public Health*, Vol. 12, pp 515.

Mathur R, Schofield P, Smith D, et al. (2017) Is individual smoking behaviour influenced by area-level ethnic density? A cross-sectional electronic health database study of inner south-east London. *ERJ Open Res*; 3: 00130-2016 [https://doi.org/10.1183/23120541.00130- 2016]

McCullagh P and Nelder JA. Generalized Linear Model, 2nd edition. NEW YORK, CHAPMAN AND HALL.

Merlo J, Wagner P, Ghith N, Leckie G (2016). An Original Stepwise Multilevel Logistic Regression Analysis of Discriminatory Accuracy: The Case of Neighborhoods and Health. *PLOS ONE,* Vol. 11, Issue 4: e0153778. doi:10.1371/journal.pone.0153778.

Moore L, Roberts C, Tudor-Smith C (2001). School smoking policies and smoking prevalence among adolescents: multilevel analysis of cross-sectional data from Wales. *Tobacco Control*, Vol.10, pp 117–123.

Noguchi M, Kobayashi T, Iwase T, Suzuki E, Kawachi I, Takao S. Social Capital and Suicidal Ideation in Community-Dwelling Older

Residents: A Multilevel Analysis of 10,094 Subjects in Japan. Am J Geriatric Psychiatry 2017; 25:37–47)

Norberg M, Lundqvist G, Nilsson M, Gilljam H, Weinehall L (2011) Changing patterns of tobacco use in a middle-aged population – the role of snus, gender, age, and education. *Glob Health Action* 2011; 4: 10.3402/gha.v4i0.5613.

Nurul HI , "Bangladesh among top smokeless tobacco using nation", http://bdnews24.com/health/2014/12/17/ Bangladesh -among-top-smokeless-tobacco-using-nations, Accessed 20 November, 2015.

O'Dwyer LM, Parker CE (2014). A primer for analyzing nested data: multilevel modeling in SPSS using an example from a REL study. *Applied Research Methods*.

Palipudi KM, Gupta PC, Sinha DN, Andes LJ, Asma S, McAfee T (2012). Social Determinants of Health and Tobacco Use in Thirteen Low and Middle Income Countries: Evidence from Global Adult Tobacco Survey. *PLoS ONE*, Vol. 7, Issue 3, e33466.

Palipudi KM, Rizwan SA, Sinha DN, Andes LJ, Amarchand R, Krishnan A, Asma S (2014). Prevalence and Sociodemographic of tobacco use in four countries of the world health organization: South-East Asia region: Finding from the Global Adult Tobacco Survey. *Indian Journal of Cancer.* Vol.51, Issue 5, Page 24-32.

Palipudi KM, Sinha DN, Choudhury S, Zaman MM, Asma S, Andes L, Dube S (2012). predictors of tobacco smoking and smokeless tobacco use among adults in Bangladesh. *Indian Journal of Cancer,* Vol. 49, Issue 4, pp 387-92. doi: 10.4103/0019-509X.107745.

Park S, Nam B-H, Yang H-R, Lee JA, Lim H, Han JT, Park Su, Shin HR, Lee JS, (2013) Individualized Risk Prediction Model for Lung Cancer in

Korean Men. *PLOS ONE* 8(2): e54823. doi:10.1371/journal.pone.0054823.

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

Piontek D, Buehler A, Donath C, Floeter S, Metz URK, Gradl S, Kroeger C (2008). School Context Variables and Students' Smoking: Testing a Mediation Model through Multilevel Analysis. *European Addiction Research*, Vol.14, pp 53–60.

Rahman MM, Karim MJ, Ahmad SA, Suhaili MR, Ahmad SNW (2011). Prevalence and Determinants of Smoking Behaviour among the Secondary School Teachers in Bangladesh. *International Journal of Public Health Research Special Issue*, pp 25-32.

Rahman MM, T areque MI, Ferdos J, Jesmin SS (2016). Maternal Pregnancy Intention and Professional Antenatal Care Utilization in Bangladesh: A Nationwide Population - Based Survey. *PLOS ONE*, Vol. 11, Issue 6: e0157760. doi:10.1371/journal.pone.0157760

Rahman MS, Akter S, Abe SK, Islam MR, Mondal MNI, Rahman JAMS, Rahman MM (2015). Awareness, Treatment, and Control of Diabetes in Bangladesh: A Nationwide Population-Based Study . *PLoS ONE* 10(2): e01 18365. doi: 10.1371/journal.pone.01 18365.

Rahman MS, Mondal NI, Islam MR, Rahman MM, Hoque MN, Alam MS (2015). Determinant factors of tobacco use among ever-married men in Bangladesh. *Drug, Healthcare and Patient Safety,* Vol. 7, pp 77-85.

Rani M, Bonu S, Jha P, Nguyen SN, Jamioum L (2003). Tobacco Use in India: prevalence and predictors of smoking and chewing in a national cross sectional household survey. *Tobacco Control;* 12:e4 (http://www.tobaccocontrol.com/cgi/content/full/12/4/e4). Access date: January 25, 2017.

Rani M, Thamarangsi T, Agarwal N. (2017) Youth tobacco use in South-East Asia: Implications for tobacco epidemic and options for its control in the region. *Indian Journal of Public Health*; Vol. 61, pp:S12-7.

Rath JM, Villanti AC, Abrams DB and Vallone DM (2012). Patterns of Tobacco Use and Dual Use in US Young Adults: The Missing Link between Youth Prevention and Adult Cessation. *Journal of Environmental and Public Health,* Volume 2012, Article ID 679134, 9 pages doi:10.1155/2012/679134.

Reda AA, Moges A, Yazew B, et al. (2012). Determinants of cigarette smoking among school adolescents in eastern Ethiopia: a cross-sectional study. *Harm Reduct J*; Vol 9, pp:39.

Reddy MM, Kanungo S, Kar SS (2018). Correlates of secondhand smoke exposure among nonsmoking youth (15–24 years) in India: Secondary analysis from Global Adult Tobacco Survey, 2009–10. *Journal of Family Medicine and Primary Care*; vol. 7, Issue 1, pp: 111-117.

Schwarz G (1978). Estimating the Dimension of a Model. *Annals of Statistics* 6: 461–464

Siahpush M, Mbiostat, Spittal M, Singh GK (2007). Association of Smoking cessation with Financial Stress and Material Well-Being: Result from a Prospective Study of a Population-Based National Survey. *American Journal of Public Health*, Vol. 97, pp 2281–2287. doi:10.2105/AJPH.2006.103580.

Singh A, Ladusingh L (2014) Prevalence and Determinants of Tobacco Use in India: Evidence from Recent Global Adult Tobacco Survey Data. *PLoS ONE* 9(12): e114073. doi:10. 1371/journal.pone.0114073

Sinha DN, Palipudi KM, Rolle I, Asma S, Rinchen S (2011). Tobacco use among youth and adults in member countries of South-East Asia region: Review of findings from surveys under the global tobacco surveillance system. *Indian Journal of Public Health*; Vol. 55, Issue 3, pp. 169-76.

Smith MI, de Lusignan S, Mullett D, Correa A, Tickner J, Jones S (2016). Predicting Falls and When to Intervene in Older People: A Multilevel Logistical Regression Model and Cost Analysis. PLoS ONE 11(7): e0159365. doi:10.1371/journal. pone.0159365

Spitz MR, Hong WK, Amos CI, WU X, Schabath MB, Dong Q, Shete S, Etzel CJ (2007). A Risk Model for Prediction of Lung Cancer. *Journal of National Cancer Institute*, Vol. 99, Issue 9, pp 715-26.

Sreeramareddy CT, Pradhan PMS, Mir IA, Sin S (2014): Smoking and smokeless tobacco use in nine South and Southeast Asian countries: prevalence estimates and social determinants from Demographic and Health Surveys. *Population Health Metrics*, Vol. 12, pp 22.

Sultana P, Akter S, Rahman MM, Alam MS (2015). Prevalence and Predictors of Current Tobacco Smoking in Bangladesh. *Journal of Biometrics and its Application*, Vol. 1, Issue 1, pp 102.

Talley B, Masyn K, Chandora R, Vivolo-Kantor A (2017). Multilevel analysis of school anti-smoking education and current cigarette use among South Africa students. *The Pan African Medical Journal, 26:37*. doi:10.11604/pamj.2017.26.37.7880

WHO, The World Health Report 2002 - Reducing Risks, Promoting Healthy Life. [Available from: http://www.who.int/whr/2002/en/ [Accessed 15 Mar 2014]

World Health Organigation (2009). Global Adult Tobacco Survey: Bangladesh Report. Dhaka, WHO Country office of Bangladesh.

World Health Organization (2008). WHO report on the global tobacco epidemic: the MPOWER package. Geneva: World Health Organization. [Available from: http://www.who.int/tobacco/mpower/mpower_report_full_2008.pdf]. Accessed on 27 April 2018.

World Health Organization, Media Center: Tobacco Fact Sheet N 339. http://www.who.int/mediacentre/factsheets/fs339/en/index.html. [Access on 04-07-2014]

Zamman MM, Nargis N (2007). Impact of Tobacco-related Illnesses in Bangladesh. ISBN 978-92-9022-290-3.