Prevalence and predictors of high nicotine dependence among adult smokers in Botswana,2017

Betelihem Getachew

4/21/23

Warning: package 'dplyr' was built under R version 4.2.2

Warning: package 'skimr' was built under R version 4.2.2

Warning: package 'png' was built under R version 4.2.2

# 1. Abstract

*Introduction:* Population level nicotine dependence research is very limited. Even more limited in low income country like Botswana. Understanding the nicotine dependence landscape in a country with low smoking prevalence country is important in designing an efficient cessation service. *Methods:* This study is a secondary data analysis utilizing data collected by Global Adult Tobacco Survey among adults. This study examined the 1)Prevalence of nicotine dependence and 2) Predictors of nicotine dependence. Results: In this sample of smokers, 46% were categorized with high addiction. Those who are men, completed secondary education, are employed, and in the lowest wealth index residing in rural areas and single marital status tend to have high nicotine addiction. *Conclusions:* Of all the predictors considered in this analysis, age of smoking initiation is the most significant predictor of nicotine dependence. *Implications:* Delayed onset of smoking initiation is associated with low nicotine dependence. Public health interventions should be aimed ensuring countries enforce policies that restrict smoking by age. In addition, almost half the smokers were high nicotine dependent, given the risk factors associated with smoking, strategies to improve access and utilization of cessations service is imperative.

# 2. Introduction

## 2.1 General Background Information

For countries like Botswana, and other African countries the prevalence of smoking is relatively low compared to other middle and high income countries. However, while the smoking prevalence might be low, not much has been said about smokers nicotine dependence in the region. Understanding nicotine dependence among smokers provides an opportunity for a targeted intervention strategies that are more efficient and effective.

Under the Global Tobacco Surveillance system Data (GTSSData) at CDC, the Global Adult Tobacco Survey (GATS) is the global standard to systematically monitor adult tobacco use and track key tobacco control indicators. The survey is a nationally representative household survey of adults 15 years of age or older, using a standard protocol. It is intended to generate comparable data within and across countries. GATS enhances countries’ capacity to design, implement and evaluate tobacco control interventions ([GATS Botswana Survey 2017](https://nccd.cdc.gov/GTSSDataSurveyResources/Ancillary/DataReports.aspx?CAID=2&Survey=4&WHORegion=3&Country=123&Site=27000)). Using GATS protocol, a nationally representative sample of 4,643 participants was collected in Botswana in 2017 using a stratified cluster sample design.

The aim of this paper is therefor to;

1. Describe the prevalence of nicotine dependence in Botswana using the Heavy Smoking Index(HSI) by Socio-demographics and other factors
2. Explore and Identify predictors of nicotine dependence in Botswana.

## 2.2 Data import and cleaning

An SPSS data file was [downloaded from GTSS Info website](https://nccd.cdc.gov/GTSSDataSurveyResources/Ancillary/DataReports.aspx?CAID=2&Survey=4&WHORegion=3&Country=123&Site=27000) and imported into R. The code for importing and cleaning the dataset is documented in the R script file titled “processingcode.R”. The dataset was examined for outliers, distribution, class appropriation. Re-catagorization of response options was conducted when appropriate.

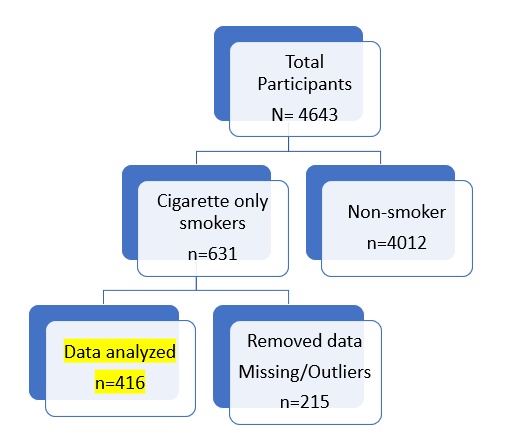
## 2.3 Methods

### 2.3.1 Study Sample Design

This study will conduct a secondary data analysis using the Global Adult Tobacco Survey (GATS) Botswana as described above. The GATS survey data collection was completed in 2017. The sample design is a multi-stage, geographically clustered probability sample design to produce nationally representative data. First, households are randomly selected, then one individual is randomly chosen from each selected household to participate in the survey. The random selection of households and participants allows for an unbiased, randomly selected, and nationally representative sample of the larger population. The cluster sampling allows representation in gender and urbanicity. More details on the Global Adult Tobacco Survey implementation process can be found elsewhere ([GATS Implementation Protocol](https://nccd.cdc.gov/GTSSDataSurveyResources/Ancillary/Documentation.aspx?SUID=4&DOCT=4)).

### 2.3.2 Participants

A total of 4,643 individuals completed the GAT’s survey. Given the aims of this study, examining nicotine addiction prevalence, non-smokers (n=4012) were removed from this data analysis. A total of 631 observations of daily and non daily smokers were retained for anlaysis. During data cleaning, 215 observations were deemed unusable data due to missing or incorrect entries and thus further removed leaving a final study sample of 416 observations.



### 2.3.3 Measurement

A total of 608 variables covering topics of socio-demographic, tobacco smoking, smokeless tobacco, cessation both smoking and smokeless, economics, media and knowledge, attitude and perceptions of tobacco were collected in this survey. Inline with the study objectives, 591 variables were filtered out and only 17 variables related to the study objectives were considered for further examination. The 18th variable was computed as a score by summing two variables (B01 Do you currently smoke daily/less than daily or not at all + B07 How soon after you wake up do you usually have your first smoke ) from the tobacco smoking section of the survey to create the Heavy Smoking Index (HSI) Score, a proxy for examining nicotine dependence ([Heaviness of Smoking Index | Data Share 2.0 (nih.gov)](https://datashare.nida.nih.gov/instrument/heaviness-of-smoking-index). Based on the HSI scores, smokers were categorized into low addiction (score 0-3), medium addiction (score 3-4), high addiction (5-6). For this analysis, in consultation with other literature, a cut off score of 4 was deemed suitable to catagorize into low/high nicotine addiction. Thus scores 0 to 3 are categorized as low nicotine addiction and scores 4 to 6 are categorized as high nicotine addiction [Heaviness of Smoking Index versus Fagerstrom Test for Nicotine Dependence among Current Smokers of Ahmedabad City, India](https://doi.org/10.22122%2Fahj.v13i1.291).

The outcome of interest therefor for this study is HSI: Heavy Smoking Index. While the Independent variables explored in this analysis include AGE: Current age of smokers, A01: Gender A04:Education level, A05: Employment status, A11: Marital Status, RESIDENCE: Residence of participants, B04: Age of smoking initiation, D01: Smoking quit attempt, D08: Smoking intent to quit.

# 3. Results

## 3.1 Descriptive analysis

In this sample of smokers, males with a median age of 35 years, those with secondary education and employed and belonging to the lowest wealth index had high nicotine addiction. There was also higher proportion of smokers who are single, live in rural for high addiction. Interms of behavior, more than half of those in the high addiction catagory made a quit attempt. However , the majority of smokers had no intention to quit within the next year . All codes related to data exploration are saved under code/analysis\_code with a file name of exploratorycodeR.R. All tables and figures listed below are saved under results folder

(**Table1\_Demographics?**) Shows the demographics of the sample population

|  | HighAddiction | LowAddiction | Overall |
| --- | --- | --- | --- |
|  | (N=192) | (N=226) | (N=418) |
| **factor(A01)** | | | |
| Female | 26 (13.5%) | 33 (14.6%) | 59 (14.1%) |
| Male | 166 (86.5%) | 193 (85.4%) | 359 (85.9%) |
| **AGE** | | | |
| Mean (SD) | 40.1 (14.2) | 40.6 (15.5) | 40.4 (14.9) |
| Median [Min, Max] | 35.0 [19.0, 77.0] | 38.0 [16.0, 97.0] | 37.0 [16.0, 97.0] |
| **B04** | | | |
| Mean (SD) | 22.4 (8.95) | 20.2 (6.54) | 21.2 (7.81) |
| Median [Min, Max] | 20.0 [1.00, 70.0] | 19.0 [5.00, 52.0] | 20.0 [1.00, 70.0] |
| **factor(A04)** | | | |
| College and above | 27 (14.1%) | 32 (14.2%) | 59 (14.1%) |
| No Formal Education | 32 (16.7%) | 48 (21.2%) | 80 (19.1%) |
| Primary Education | 37 (19.3%) | 43 (19.0%) | 80 (19.1%) |
| Secondary Education | 96 (50.0%) | 101 (44.7%) | 197 (47.1%) |
| Missing | 0 (0%) | 2 (0.9%) | 2 (0.5%) |
| **factor(A05)** | | | |
| Employed | 108 (56.3%) | 122 (54.0%) | 230 (55.0%) |
| Unemployed | 84 (43.8%) | 104 (46.0%) | 188 (45.0%) |
| **factor(Wealth)** | | | |
| High | 21 (10.9%) | 28 (12.4%) | 49 (11.7%) |
| Higher | 34 (17.7%) | 40 (17.7%) | 74 (17.7%) |
| Low | 17 (8.9%) | 16 (7.1%) | 33 (7.9%) |
| Lowest | 86 (44.8%) | 99 (43.8%) | 185 (44.3%) |
| Middle | 34 (17.7%) | 43 (19.0%) | 77 (18.4%) |
| **factor(A11)** | | | |
| Married | 21 (10.9%) | 34 (15.0%) | 55 (13.2%) |
| Single | 171 (89.1%) | 192 (85.0%) | 363 (86.8%) |
| **factor(RESIDENCE)** | | | |
| Rural | 109 (56.8%) | 123 (54.4%) | 232 (55.5%) |
| Urban | 83 (43.2%) | 103 (45.6%) | 186 (44.5%) |
| **factor(D01)** | | | |
| No | 85 (44.3%) | 111 (49.1%) | 196 (46.9%) |
| Yes | 107 (55.7%) | 115 (50.9%) | 222 (53.1%) |
| **factor(D08)** | | | |
| No | 128 (66.7%) | 173 (76.5%) | 301 (72.0%) |
| Yes | 63 (32.8%) | 53 (23.5%) | 116 (27.8%) |
| Missing | 1 (0.5%) | 0 (0%) | 1 (0.2%) |

The data also shows the mean age of those with high addiction to be lower than the mean age of those with low addiction. This is contrary to the belief that older smokers have higher addiction when compared to younger smokers.

(**figure1?**) shows a boxplot showing HSI Group mean by age

|  |
| --- |
| HSI By Age. |

Interms of age of smoking initiation, those in the high addiction group have a mean age of smoking initiation slightly higher than those in the low addiction group. This again is contrary to the literature that suggests starting smoking early leads to high addiction in adulthood.

(**figure2?**) shows a boxplot showing HSI Group mean age of smoking initiation

|  |
| --- |
| HSI By Age of smoking Initiation. |

## 3.2 Basic statistical analysis

## 3.3 Full analysis

A binary 5-fold cross validation logistic regression was performed. Model I used all predictors, Model II used only the age of initiation as the predictor and model III used cessation behaviors (quit attempt ) as predictors. The tables below represent the results for all three models

(**Model1?**) All predictors

# A tibble: 16 × 5  
 term estimate std.error statistic p.value  
 <chr> <dbl> <dbl> <dbl> <dbl>  
 1 (Intercept) 1.69 0.940 1.80 0.0725   
 2 AGE 0.00131 0.0114 0.114 0.909   
 3 B04 -0.0459 0.0174 -2.63 0.00842  
 4 A01\_Male -0.0192 0.352 -0.0547 0.956   
 5 RESIDENCE\_Urban 0.170 0.264 0.642 0.521   
 6 A04\_No.Formal.Education 0.178 0.533 0.334 0.739   
 7 A04\_Primary.Education 0.254 0.489 0.521 0.603   
 8 A04\_Secondary.Education -0.144 0.387 -0.371 0.710   
 9 A05\_Unemployed 0.174 0.239 0.728 0.467   
10 A11\_Single -0.432 0.394 -1.10 0.273   
11 Wealth\_Higher -0.403 0.446 -0.903 0.367   
12 Wealth\_Low -0.321 0.536 -0.599 0.549   
13 Wealth\_Lowest -0.377 0.414 -0.913 0.361   
14 Wealth\_Middle 0.0380 0.443 0.0856 0.932   
15 D08\_Yes -0.167 0.270 -0.619 0.536   
16 D01\_Yes -0.136 0.245 -0.555 0.579

(**Model2?**) Age of initiation as predictor

# A tibble: 2 × 5  
 term estimate std.error statistic p.value  
 <chr> <dbl> <dbl> <dbl> <dbl>  
1 (Intercept) 0.990 0.350 2.83 0.00472  
2 B04 -0.0390 0.0157 -2.49 0.0128

(**Model3?**) Quit attempt as predictors

# A tibble: 2 × 5  
 term estimate std.error statistic p.value  
 <chr> <dbl> <dbl> <dbl> <dbl>  
1 (Intercept) 0.267 0.169 1.58 0.113  
2 D01\_Yes -0.196 0.229 -0.856 0.392

### 3.3.1 Model Evaluation and Performance using CV folds

The three models were evaluated based on accuracy and roc\_auc scores. Compared to model I and model III, model II performed better and thus selected for final test fit. The metrics for all three model evaluations considered are presented below.

Model I Performance

# A tibble: 2 × 6  
 .metric .estimator mean n std\_err .config   
 <chr> <chr> <dbl> <int> <dbl> <chr>   
1 accuracy binary 0.524 5 0.0239 Preprocessor1\_Model1  
2 roc\_auc binary 0.510 5 0.0305 Preprocessor1\_Model1

Model II Performance

# A tibble: 2 × 6  
 .metric .estimator mean n std\_err .config   
 <chr> <chr> <dbl> <int> <dbl> <chr>   
1 accuracy binary 0.563 5 0.0350 Preprocessor1\_Model1  
2 roc\_auc binary 0.599 5 0.0178 Preprocessor1\_Model1

Model III performance

# A tibble: 2 × 6  
 .metric .estimator mean n std\_err .config   
 <chr> <chr> <dbl> <int> <dbl> <chr>   
1 accuracy binary 0.518 5 0.0395 Preprocessor1\_Model1  
2 roc\_auc binary 0.527 5 0.0348 Preprocessor1\_Model1

### 3.3.2 Final Model

|  |
| --- |
| Predictive Model |

# 4. Discussion

## 4.1 Summary and Interpretation

All three models used in this anlaysis had an roc\_auc estimate of .59 or less. This means these models do not do a good job of predicting addiction with any more accuracy than a guess with a 50/50 outcome except for model II. This also implies further literature review is required to narrow down some of the most notable predictors. Based on this analysis , age of smoking initiation is a statistically significant predictor of nicotine dependence.

## 4.2 Strengths and Limitations

*Discuss what you perceive as strengths and limitations of your analysis.*

## 4.3 Conclusions

*What are the main take-home messages?*

*Include citations in your Rmd file using bibtex, the list of references will automatically be placed at the end*

This paper (Leek & Peng, 2015) discusses types of analyses.

These papers (McKay, Ebell, Billings, et al., 2020; McKay, Ebell, Dale, Shen, & Handel, 2020) are good examples of papers published using a fully reproducible setup similar to the one shown in this template.

Note that this cited reference will show up at the end of the document, the reference formatting is determined by the CSL file specified in the YAML header. Many more style files for almost any journal [are available](https://www.zotero.org/styles). You also specify the location of your bibtex reference file in the YAML. You can call your reference file anything you like, I just used the generic word references.bib but giving it a more descriptive name is probably better.

# 5. References

Leek, J. T., & Peng, R. D. (2015). Statistics. What is the question? *Science (New York, N.Y.)*, *347*(6228), 1314–1315. <https://doi.org/10.1126/science.aaa6146>

McKay, B., Ebell, M., Billings, W. Z., Dale, A. P., Shen, Y., & Handel, A. (2020). Associations Between Relative Viral Load at Diagnosis and Influenza A Symptoms and Recovery. *Open Forum Infectious Diseases*, *7*(11), ofaa494. <https://doi.org/10.1093/ofid/ofaa494>

McKay, B., Ebell, M., Dale, A. P., Shen, Y., & Handel, A. (2020). Virulence-mediated infectiousness and activity trade-offs and their impact on transmission potential of influenza patients. *Proceedings. Biological Sciences*, *287*(1927), 20200496. <https://doi.org/10.1098/rspb.2020.0496>