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

Betelihem Getachew

2/24/23

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

# 1. Introduction

## 1.1 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.

To cite other work (important everywhere, but likely happens first in introduction), make sure your references are in the bibtex file specified in the YAML header above (here dataanalysis\_template\_references.bib) and have the right bibtex key. Then you can include like this:

Examples of reproducible research projects can for instance be found in (McKay, Ebell, Billings, et al., 2020; McKay, Ebell, Dale, Shen, & Handel, 2020)

# 2. Methods

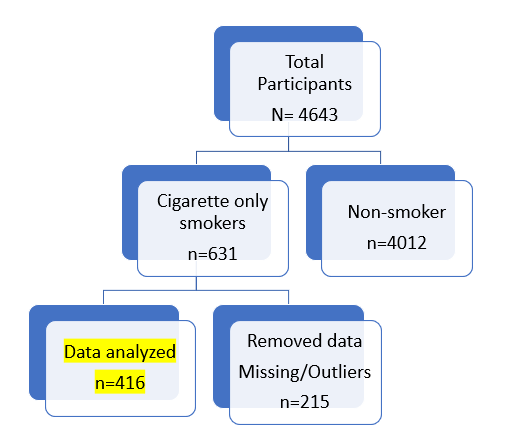
*Describe your methods. That should describe the data, the cleaning processes, and the analysis approaches. You might want to provide a shorter description here and all the details in the supplement.*

This study will conduct a secondary data analysis using the Global Adult Tobacco Survey (GATS) Botswana as described above. The 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)).

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. While the outcome measure of interest “Nicotine dependence” is not directly collected, the measure for nicotine dependence, also known as the Heavy Smoking Index (HSI), will be calculated as a score using two question items from the tobacco smoking section on the survey questionnaire ([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 will be categorized into low addiction (score 0-2), medium addiction (score 3-4), high addiction (5-6).

## 2.1 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 raw dataset contained 4643 observations and 608 variables, of which 591 variables were filtered out. Only 17 variables were considered for further examination.The variables of interest included socio-demographics, smoking status, smoking behaviors, cessation and media exposure. The 18th variable was computed by summing two variables (B01 + B07) to create the Heavy Smoking Index (HSI) Score, a proxy for examining nicotine dependence. The number of observations was further reduced by including only daily and less than daily smokers leaving a total of 631 observations. The dataset was examined for outliers, distribution, class appropriation. Recatagorization of response options was conducted when appropriate. After the removal of missing and outlier values, a total of 416 observations and 18 variables were saved for data exploration.



—- exploredata ——–

## 2.2 Statistical analysis

*Explain anything related to your statistical analyses.*

# 3. Results

## 3.1 Exploratory/Descriptive analysis

In this sample of smokers, a higher proportion of smokers who completed secondary education were categorized in all three addiction levels. Married individuals with Higher Wealth level were also catagorized in different level the highest proportion being under high addiction. In all three catagories of HSI, participants of rural residence were larger in proportion than urban residents.

(**table1?**) shows a summary of the data.

(**figure1?**) shows boxplots

(**figure2?**) shows histogram

## 3.2 Basic statistical analysis

*To get some further insight into your data, if reasonable you could compute simple statistics (e.g. simple models with 1 predictor) to look for associations between your outcome(s) and each individual predictor variable. Though note that unless you pre-specified the outcome and main exposure, any “p<0.05 means statistical significance” interpretation is not valid.*

[Figure 1](#fig-result) shows a scatterplot figure produced by one of the R scripts.

|  |
| --- |
| Figure 1: Height and weight stratified by gender. |

## 3.3 Full analysis

*Use one or several suitable statistical/machine learning methods to analyze your data and to produce meaningful figures, tables, etc. This might again be code that is best placed in one or several separate R scripts that need to be well documented. You want the code to produce figures and data ready for display as tables, and save those. Then you load them here.*

Example [Table 1](#tbl-resulttable2) shows a summary of a linear model fit.

Table 1: Linear model fit table.

| term | estimate | std.error | statistic | p.value |
| --- | --- | --- | --- | --- |
| (Intercept) | 149.2726967 | 23.3823360 | 6.3839942 | 0.0013962 |
| Weight | 0.2623972 | 0.3512436 | 0.7470519 | 0.4886517 |
| GenderM | -2.1244913 | 15.5488953 | -0.1366329 | 0.8966520 |
| GenderO | -4.7644739 | 19.0114155 | -0.2506112 | 0.8120871 |

# 4. Discussion

## 4.1 Summary and Interpretation

*Summarize what you did, what you found and what it means.*

## 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, et al., 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>