Contents

Part I
Preliminaries

1 Introduction

Tobacco, development, equity

The main goal of the grant is to deepen and expand the use of extended cost-effectiveness analyses (ECEA) to provide additional analytic support for local taxation decisions in Colombia (C), India (I) and Mexico (M) [1]. The project has the following specific objectives[2]:

1. Synchronize local expertise and strengthen research capacity on tobacco economics in each country, aligned with the Country Coordinating Mechanism - CCM. (MIC)

2. Use ECEA to investigate the national and subnational health and financial protection benefits that are associated with significant excise tax increases in each country and compare how these benefits differ by subnational area and by income groups and sex in each country and extend the use of ECEA by making it available in an open source version. (MIC)

3. Deepen and extend the ECEA framework and develop ECEA2 which will be a model that can account for the household costs associated with tobacco consumption over the life-course, gender and family effects, as well as more accurately account for declining levels of tobacco consumption over time. (C)

4. To generate tax diagnostics for each country, using evidence from the ECEAs, and implement them as policy tools to inform decision-making during the planning cycle in each country. (MIC)

5. To determine the factors that influence the uptake and use of data on the impact of fiscal and economic strategies by policy makers in each country. (MI)

This document presents the subnational analysis for Colombia (objective 2) as well as the open source version of the ECEA for tobacco tax policies.

Strengthen CCM in Each Country.... First, we will undertake a careful mapping of all tobacco taxation research, advocacy and policy initiatives in each country. ... It will also inform the selection of subnational areas with the greatest potential for successfully implementing and benefiting from tobacco taxation. [1, p.9]

1 [1, p.3]
2 [1, p.27]

Part II
Conceptual framework

2 Development: Income and equity - NM

3 ECEA - NM

One of the main contributions of this exercise is to take advantage of the flexibility of the ECEA to model the particular conditions of the country. Since the analysis is not for international comparison but for policy discussion in the country, the ECEA can be modified to follow closer the particular conditions of the country in terms on data inputs as well as in modeling (e.g. the structure of the tax)

4 Subnational

In Colombia,

Figure 1: Departments in Colombia (33)

Figure 2: Departments in Colombia (33)

Figure 3: Regions in Colombia (5)

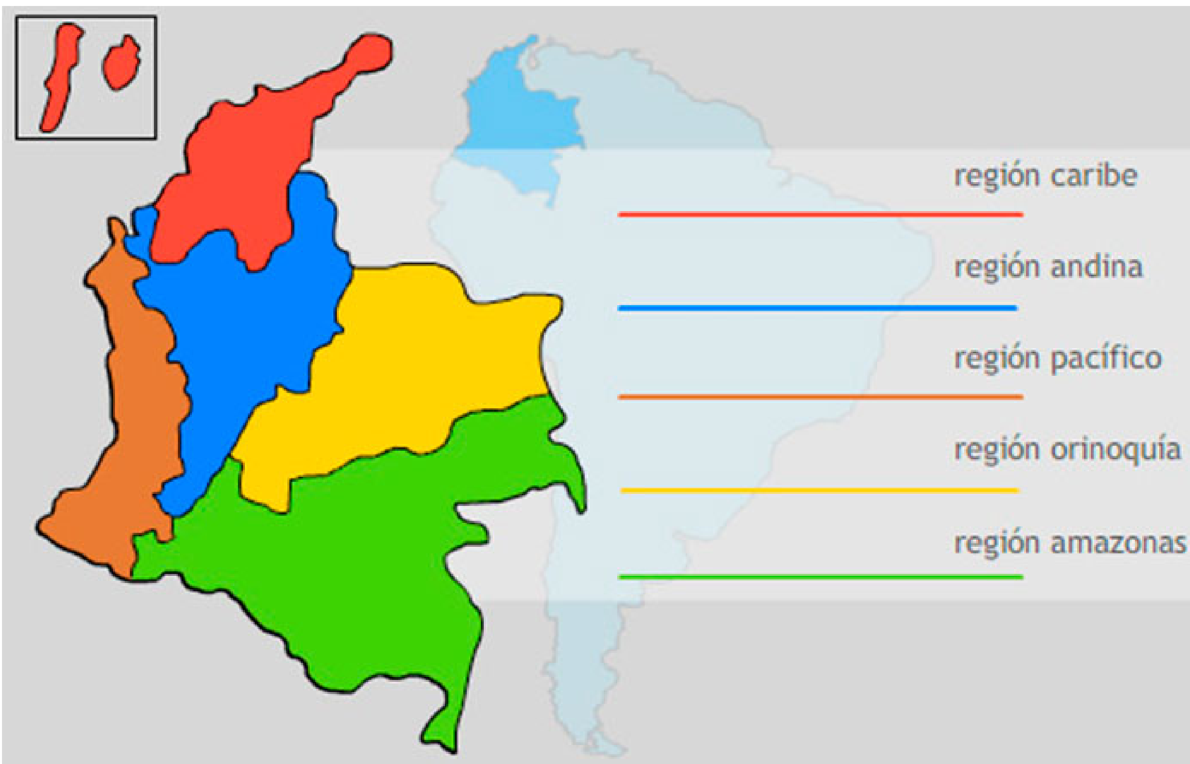
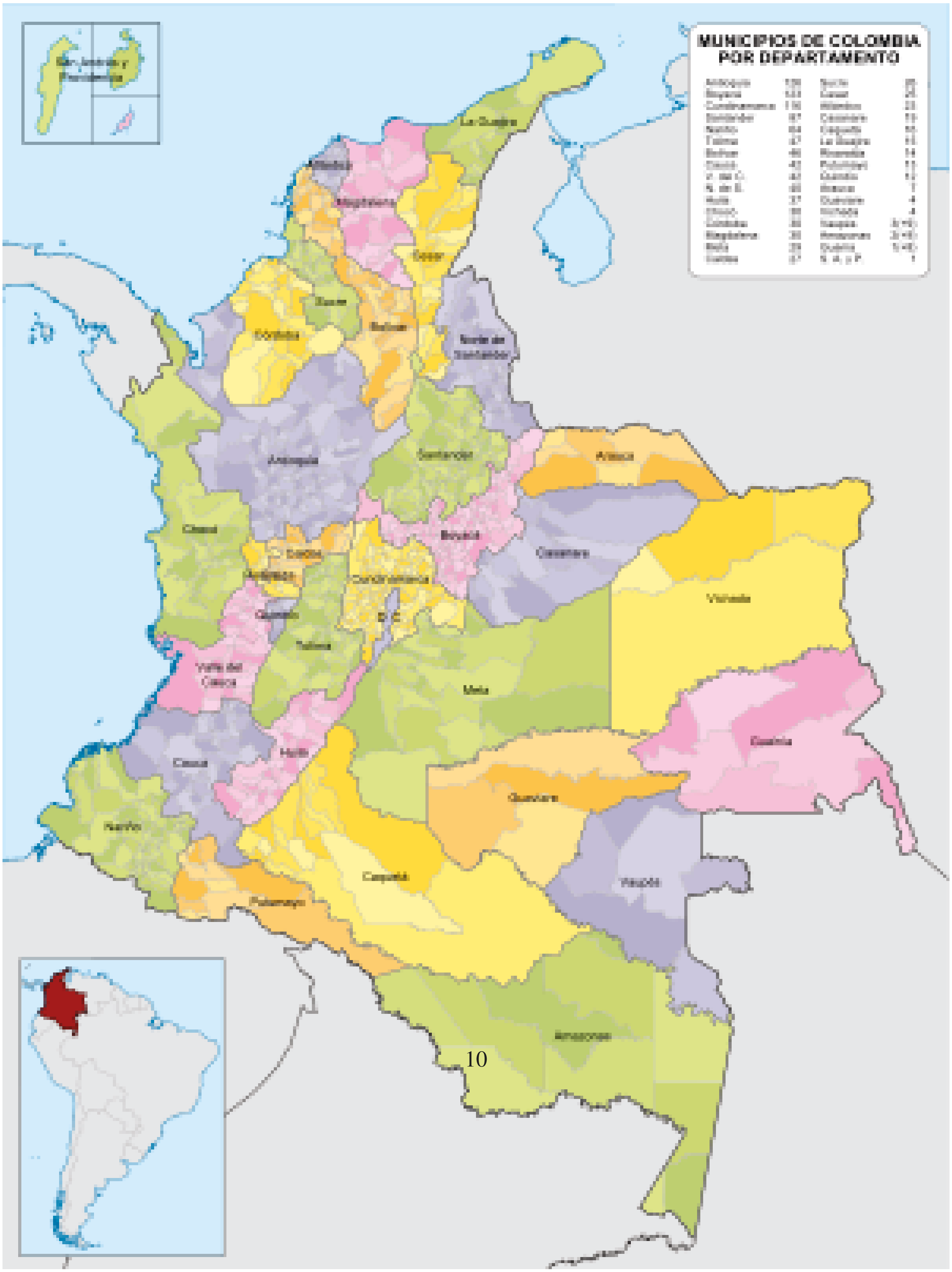


Figure 4: Municipios in Colombia (1122)



4.1 Relevance for policy in Colombia - NM

4.2 Definition of regions - SH

5 Decision makers and ECEA outputs (messages) - BLL

5.1 Colombia

5.1.1 Tobacco tax

5.1.2 Agents: National

Ministry of Health

Ministry of Finance

Congress

5.1.3 Subnational

Departamentos and Governors

Municipalities and Majors

Regions

5.1.4 National and subnational

Part III
Method

6 ECEA

The method to carry out ECEA is developed in [2] and the praxis is presented in [3]. The model follows Verguet [3] and the application of the ECEA analysis for tobacco taxes in Colombia [4, 5]. The steps of the ECEA for tobacco tax policy are as follows:

Step 1. Select the policy instrument. The health policy (HP) for this analysis is excise taxes on tobacco. To simplify the analysis, the focus is on the specific component of the excise tax; this simplification avoids the complications associated to simultaneity between prices and advalorem taxes. Also, it is recommended to focus efforts on the specific component because, compared to the advalorem component of the exices tax, the specific component cannot be manipulated with pricing strategies from the industry.

Step 2. Define target population and population segmentation. The target population is the population in Colombia. Segmentation of the population P is associated with income quintiles, meaning population is divided into 5 subgroups (i.e. $P_k \ \forall \ k = 1, \ldots, 5$).

Step 3. Obtain parameters. Parameters are chosen from literature or estimated using available data sources. Simulations are also performed when necessary. For the colombian case, enough data is found to make income simulations unnecessary.

Step 4. Build indicators. Define and build indicators using results obtained. These indicators proof useful to obtaining conclusions, further work and policy recommendations.

The model is divided in 6 modules. The first module contains information regarding the epidemic (e.g. tobacco prevalence). The second module concerns information on tobacco consumption. A third module is focused on the health variables included in the ECEA. Following health, the fourth module shows effects in healthcare (health costs and utilization, among others). A fifth module presents the information regarding tax revenue. Lastly, the sixth module is concentrated on the effects in poverty and averted poverty.

From now and forth, the structure of the model will be described in detail, going through each module indicating data source and method for either original BMJ work and GTEC-Subnational. It is worth noting that national outcomes result from aggregating all the subnational calculations, this is done to satisfy the macro-consistency (the sum of subnational equals the national). The statistical measure which allows to disaggregate national to subnational data was the coefficient of variation. Acceptable estimates were those whose coefficient of variation was less than 30% [6, p.170] and formulas used were according to the type of variable [7, p.60]. All input values used in the BMJ paper are available in the Excel spreadsheet *ECEA_stata_v1.csv*.

6.1 Epidemic

The epidemic module is presented as the first module. Its place as the first module matches its importance in the ECEA model as a whole. In this module, key parameters and variables are defined which determine almost all the other module's results. The parameters and variables included in this module are: price elasticities, tobacco prevalence, population, risk reduction, post-tax proportion of adult and young smokers, and the percent change in price (in absolute value). The next subsections present each of these parameters and variables, presenting its calculations in BMJ as well as GTEC-Subnational.

6.1.1 Price elasticities

Price elasticities correspond to the percent change in quantities due to a change in prices. It can be useful for quitting rates calculations. Price elasticities are calculated for the different income quintiles in both BMJ and GTEC-Sub.

1. Source.

   (a) BMJ. In Global Tobacco Economics Consortium, the elasticity is -0.4 [8, p.2], using estimates from [9, 10]. However, in the supplementary appendix the value for Colombia is -0.78 [8, Supplementary appendix, p.4]. In addition, the Stata code for the paper has an elasticity of -0.635 for the lowest quintile and -0.122 for the highest quintile.

   (b) GTEC-Sub. The most recent estimates with disaggregation by income are the ones estimated in a working paper [11]. These estimates correspond to smoking prevalence elasticities for 3 socioeconomic groups (strata): strata 1-2: -0.222, stratum 3: -0.228 and strata 4-6: -0.232. Another estimate was obtained using aggregated quarterly data for the period 1994-2014, resulting in a price elasticity of -0.78 [12]. Other calculations found an aggregated demand price elasticity ranging from -0.3 to -0.42 [13, 14]. Within a study driven for Colombia using an extended cost-effectiveness analysis methodology, price elasticities per income quintile were calculated. Using an elasticity based on household data (ECV for 2003, 2010, 2011, 2014) of -0.44 for Colombia (PE_{Col}), an elasticity mean of -0.53 (m) and interquartile range of 0.42 (iqr) resulted from 11 studies carried out for 8 countries in latinamerica, an interpolation was implemented to derive price elasticity per income quintile in Colombia [15, 16]. The latter procedure is detailed in the following equations:

$$PE_1 = PE_{Col} + \frac{iqr}{2m} PE_{Col} = -0.61$$
$$PE_2 = PE_{Col} + \frac{iqr}{4m} PE_{Col} = -0.53$$

$$PE_3 = PE_{Col} = -0.44$$

$$PE_4 = PE_{Col} - \frac{iqr}{4m} PE_{Col} = -0.35$$
$$PE_5 = PE_{Col} - \frac{iqr}{2m} PE_{Col} = -0.26$$

Where PE_i, is the price elasticity of quintile i and the elasticity for Colombia (-0.44) when $i = Col$, and iqr and m are the interquartile range and the mean as discussed above.

2. Method:

   (a) BMJ/GTEC-Sub. We used the same quintile elasticities (elq) for Colombia as the Stata code did for the BMJ paper, that is, -0.635 and -0.122 for the lowest and highest quintiles, respectively. For the other quintiles, a linear interpolation is done between the elasticity of the highest quintile and the elasticity of the lowest one.

6.1.2 Prevalence

Tobacco prevalence is a key parameter in the whole ECEA model. Tobacco prevalence is estimated for each income quintile and 5-years age group. Once prevalences are estimated, relative prevalences in each quintile are calculated from the mean prevalence in all quintiles. The process and data sources are described below.

1. Source.

   (a) BMJ. For the BMJ paper, prevalence is calculated using data available in ECSP 2013 [17]. Prevalence by income quintile is estimated using education level as a proxy for income. No education was taken as the first quintile, primary for the second, middle school for the third, high school for the fourth and college for the fifth. We validated these data by comparing values in the BMJ supplementary appendix [8, Supplementary appendix], values in the Stata code and values in the cited source [17].

   (b) GTEC-Sub. Prevalence in each of the 6 regions as well as the one for the country were obtained by combining two sources of information: Encuesta de Calidad de Vida 2018 - ECV2018 (Quality of Life Survey) y Encuesta de Consumo de Sustancias Psicoactivas en Poblacion Escolar 2016 - ECSPPE2016 (National Psychoactive Substances Survey in Scholar Population). Using ECV2018, current smoking prevalence was computed per income quintile stratifying by *percapita* variable, which measures the level of income of the expenditure unit (usually the household).

To calculate prevalence for every age group ($smokingprev\_age$), for adults (people aged 18 or older) we calculated current smoking prevalence from ECV2018 stratifying by income quintiles using the variable *percapita*, which measures the level of income of the expenditure unit (usually the household). For ages below 18, ECV2018 was dismissed because ECV18 asks about smoking in front of all household members, so, there is an incentive for young people to hide information on smoking habits from their parents, leading to a significant underreport of smoking in young population in ECV2018. For this reason, we use ECSPPE2016 to calculate prevalences for this segment of population as this survey uses a protocol to interview individually. We calculated the prevalence for 12-14 and 15-18 ranges and used these estimates as prevalence for groups 10-14 and 15-19 in the model, respectively.

Since this survey does not have a variable for income, we used education as a proxy for income. To do this, we transformed highest level of education reached by each parent into number of years studied. The imputation in each level was the following: no studies was taken as 0 years; primary took a value of 5 years; high school correspond to 11 years; technician was 12 years worth; undergraduate took 16 years; and postgraduate was taken as 18 years. Total years studied per household equals the sum of years studied by each parent. Using the total number of education years, quintiles were computed.

2. Method.

   (a) BMJ/GTEC-Sub. Although processes taken in each data source to obtain prevalences are different between BMJ and GTEC-Sub, the calculation is the same in both. Finally, relative prevalences to the mean ($prev$) of the 5 prevalences by education level are calculated.

$$\begin{aligned} prev_q &= \frac{prev_q}{\frac{\sum prev_q}{5}} \\ &= \frac{prev_q}{0.2 * \sum prev_q} \end{aligned}$$

   Where:

$$prev_q = \frac{smokers_q}{pop_q}$$

6.1.3 Population

Population at a national and subnational level are calculated. At each level, population in 5-age groups up to a 100 years is used in each model. Population is necessary to calculate changes in number of smokers (using prevalences) and thus, changes in all other related variables.

1. Source

    (a) BMJ. Population in 5-year age group ($popage$) was obtained from United Nations population estimates and projections for Colombia [18]. Data used correspond to 2015 census projections and can be seen in BMJ paper supplement [8, Supplementary appendiz, p.3].

    (b) GTEC-Sub. Population estimates from Colombia's National Statistical System for 2018 are taken as the source for population in 5-year age groups for the GTEC-Sub. When compared to the most recent census in 2018, population from 2005 census' projections is overestimated. (e.g. for Bogota, overestimation is 999,578). Moreover, while checking wheter macro-consistency was achieved, using the 2005 census projections led to no consistency. This is due to differences at the subnational distribution of population between Encuesta de Calidad de Vida and the 2005 census population projections in 2018. As samples of national surveys such as Encuesta de Calidad de Vida were calculated based on the 2005 census and since we are using those surveys we decided to stick with projections from the 2005 census population.

2. Method

    (a) BMJ. Authors take population projections as presented from the UN population estimates. This estimates are built upon determining the best method between a Modified Lee-Carter (MLC) and a Pattern of Mortality Decline (PMC) depending on the country's data availability.

    (b) GTEC-Sub. Population ($popage$) for the GTEC-Sub is taken from the population projections developed by Colombia's National Statistical System for the year 2018. Subnational distributions by 5-age groups are built by aggregating the different groups at each subnational level. Population projections are aggregated after 80 years while 5-year age groups in the ECEA model are up to 100 years. Thus, a final step consists in applying the same proportions for Colombia available in the ECEA developed by BMJ to the last aggregated age group in the colombian population projections at each subnational level. National results are obtained by aggregating the calculated subnational population to guarantee macro-consistency among population.

6.1.4 Risk reduction

A risk reduction due to quitting is taken into account for each age-group at quitting. This risk reduction is used to calculate averted deaths due to smoking. There is no source for Colombia regarding risk reduction by age, thus risk reductions used in BMJ are used for GTEC-Sub as well.

1. Source

(a) BMJ. BMJ paper authors took risk reduction factor ($riskreduction$) by 5-year age group from prior calculations made in a similar study conducted for China. Values can be checked in the supplementary appendix of this paper [19].

(b) GTEC-Sub. We adopted the same risk reduction values as BMJ did for Colombia's calculations consigned in the Stata code.

2. Method

   (a) BMJ. Based on age-specific relative risk reduction estimated for United Kingdom [20], BMJ authors used a cubic splines to compute 5-year age group factors.

6.1.5 Quitting

Quitting is determined using quit rates for each quintile. Quitting rates are determined from price elasticity, the price increase and an impact from the price increase in quitting decision. The impact is defined as halve the effect from the price increase towards quitting decision and another halve towards tobacco consumption (this is taken from the initial BMJ model).

1. Method

   (a) BMJ/GTEC-Sub. Generating quit rates ($quit$) for each income group based on elasticity, increase and impact of price increase.

$$quit_q = -el_q * imp\_pr * pr\_inc$$

6.1.6 Non-quitting adult smokers

Only a proportion of current smokers in the pre-tax scenario will quit due to the price increase. The proportion of smokers who keep smoking after the tax is used to determine revenue in the post-tax scenario, as well as health outcomes and expenditures which further lead to poverty.

1. Method

   (a) BMJ/GTEC-Sub. Share of smokers who do not quit ($consq$) is calculated based on elasticity, the 50% price increase and its impact on quitting.

$$cons_q = \frac{1 + el_q * pr\_inc}{1 + el_q * pr\_inc * imp\_pr}$$

6.1.7 Non-quitting young smokers

As it occurs in adults, in young people only a proportion of current smokers in the pre-tax scenario will quit due to the price increase. The calculation is almost te same as in adults but the expected effect is twice as in adults.

1. Method

   (a) BMJ/GTEC-Sub. Similar procedure as before but for youth. Now the twice effect in this population is included ($consqyouth$).

$$consyouth_q = \frac{1 + (2 * el_q * pr\_inc)}{1 + (2 * el_q * pr\_inc * imp\_pr)}$$

6.1.8 Switch

Finally, the change in consumption due to the simulated change in price (i.e. the elasticity times the percentage change in price) in absolute value is presented in the model.

1. Method

   (a) BMJ/GTEC-Sub.

$$switch_q = -el_q * pr\_inc$$

6.2 Consumption

The consumption module presents the variables related to the number of smokers among young and adult population, as well as pre-tax and post-tax number of smokers. The module also presents the total number of smokers in both, pre-tax and post-tax scenarios.

6.2.1 Baseline adult smokers

Baseline adult smokers is the number of adult smokers in the pre-tax scenario. This calculation is performed by calculating the number of smokers by age-group and income quintile using the calculated population by age-group and prevalence in each quintile. The resulting number of smokers is aggregated to obtain the final result by income quintile.

1. Method

   (a) BMJ/GTEC-Sub. For the calculation of baseline adult smokers by income group ($adult\_tot\_blsmkr$), several intermediate steps took place. The procedure is detailed next:

i. Prevalences for each age group jointly with its total population were computed to obtain the number of baseline smokers in each age group i.

$$bl\_smkr_i = prev_i * pop_i$$

ii. This number of smokers per age groups is splitted into income quintiles.

$$bl\_smkr_{iq} = bl\_smkr_i * prev_q * 0.2$$

The existence of 0.2 number is to find the single income quintile share of total smokers. It is, what amount of total smokers corresponds to each quintile. Thus, the 0.2 permits to transform the relative-to-mean prevalence into a proportion of prevalences, which is the suitable measure to perform number of smokers in each quintile for every age group.

iii. From this smokers, filtrate for adults only (>15 years) and sum for age group i > 3 in order to get adult smokers per income quintile.

$$adult\_tot\_blsmkr_q = \sum_{i=4}^{21} bl\_smkr_{iq}$$

6.2.2 Potential young smokers

Potential young smokers follow the same steps as the baseline adult smokers. The difference is among the age in which young smokers are considered individuals 15 years or younger.

1. Method

   (a) BMJ/GTEC-Sub. Total number of potential young smokers (tot_blsmkr_youth) at baseline (less than 15 years). Smokers not considered before are summed per income quintiles, it is people less than 15 years (age groups i <= 3)

$$tot\_blsmkr\_youth_q = \sum_{i=1}^{3} bl\_smkr_{iq}$$

6.2.3 Total smokers

For each quintile the total number of smokers is calculated aggregating the number of smokers among the 5-year groups in each quintile.

1. Method

(a) BMJ/GTEC-Sub. Not discriminating for age range, for each income quintile, sum for every age group i to obtain the total number of smokers (tot_blsmkr).

$$tot\_smkr_q = \sum_{i=1}^{21} bl\_smkr_{iq}$$

6.2.4 Adult post smokers

Post smokers in each quintile are calculated as the total number of smokers who remain smoking. This quantity is equal to the number of smokers times the complement of the quitting rate. For the young population quitting rates have double the effect.

1. Method

   (a) BMJ/GTEC-Sub. In order to obtain adult post smokers by income quintiles ($adult\_totpostsmkr$), prior calculations are taken.

      i. First, if the age group is younger than 20 years, the "remain" rate ($1 - quitrate_q$) is lower for each age group and quintile as the quit rate is multiplied by 2. The result is further multiplied by the corresponding number of baseline smokers to get the result.

      $$postsmkrs_{iq} = bl\_smkr_{iq} * (1 - quitrate_q * 2)$$

      ii. If the age group is 20 years or older, the same procedure applies as before but excluding the young effect in quitting rates.

      $$postsmkrs_{iq} = bl\_smkr_{iq} * (1 - quitrate_q)$$

      iii. The total number of adult post smokers is then the sum of post smokers aged 15 years or older (age group 4 or higher).

      $$adult\_totpotsmkrs_q = \sum_{i=4}^{21} postsmkr_{iq}$$

6.2.5 Total post smokers

The total number of post smokers is the total number of adult post smokers plus the total number of young post smokers.

1. Method

(a) BMJ/GTEC-Sub. Including youngers, for each q, totalize the number of post smokers ($tot\_postsmkr$) in every age group i.

$$tot\_postsmkr_q = \sum_{i=1}^{21} postsmkr_{iq}$$

6.2.6 Smokers prevented from initiation

Total smokers prevented are calculated as the total number of former smokers. Smokers prevented from initiation are estimated as the total number of former smokers aged 24 years or lower.

1. Method

   (a) BMJ/GTEC-Sub. Smokers prevented ($tot\_fmrsmkr$) are calculated as smokers prevented from initiation (ages 0-15 years) plus former smokers in ages 15 to 24 years. The next steps are taken to perform this calculation:

   i. If the age group is younger than 20 years, the baseline smokers in age group i and quintile q is multiplied by twice the $quitrate_q$ (twice the effect in younger population).

   $$fmrsmkr_{iq} = bl\_smkr_{iq} * quitrate_q * 2$$

   ii. If the age group is 20 years or older, baseline smokers in age group i and income quintile q is multiplied by the $quitrate_q$.

   $$fmrsmkr_{iq} = bl\_smkr_{iq} * quitrate_q$$

   iii. Finally the sum prevented smokers in the entire population for every income quintile q result in the total number of prevented smokers.

   $$tot\_fmrsmkr_q = \sum_{i=1}^{21} fmrsmkr_{iq}$$

6.3 Health

The health module includes information regarding burden due to tobacco consumption in general, as well as, burden due to 4 major diseases related to tobacco consumption. In this sense, years of life gained and deaths averted are calculated and presented in this module.

6.3.1 Years life gained

As an effect from quitting smoking, individuals live longer. The years they get to live longer due to quitting are referred as years of life gained. This are estimated using epidemiological information and population by age group.

1. Source

   (a) BMJ. Number of life years gained from quitting smoking in an specific age range were obtained from a study conducted for United States using information from 1997 to 2004. For smokers who quit at 25 to 34, 35 to 44, 45 to 54 and 55 to 64 years of age, the life years gained compare to those who continued smoking were 10, 9, 6 and 4, respectively [21]. BMJ authors used these numbers for specific ages, as follows: 10, 9, 6 and 4 years gained for 15, 25, 45 and 65 year olds, respectively.

   (b) GTEC-Sub. We adopted the same numbers as BMJ authors did in the Stata code for Colombia's calculations.

2. Method

   (a) BMJ/GTEC-Sub. Years life gained ($totallyg$) are estimated by income quintiles. To do so, several steps took place before, its process is described below.

      i. From these specific-age life years gained, a linear interpolation was computed in order to assign a single value to every age considered in the model (0-105 years).

      ii. Grouping specific ages into 5-years age groups, the average of years gained is computed. So, ages ranging from 1 to 4, receive the value of 1, 5 to 9 are classified in group 2, 10 to 14 in group 3 and so on. Then, average if found for every of the 21 of the clusters.

      iii. The smokers prevented for initiation in every age group i and income quintile q is multiplied by its correspondent average years gained calculated in previous step.

$$lyg_{iq} = fmrsmkr_{iq} * \overline{ly}_i$$

      iv. Finally, across age groups, life years gained are summed by income quintiles.

$$totallyg_q = \sum_{i=1}^{21} lyg_{iq}$$

6.3.2 Smokers' deaths averted

Smokers' deaths averted are the number of deaths that can be attributed to tobacco taking into account the risk reduction from quitting smoking, which reduces the amount of deaths that can be attributed to tobacco conssumption.

1. Method

   (a) BMJ/GTEC-Sub.To calculate deaths averted ($totfrmrdhs$) the following steps took place.

      i. Multiply smokers prevented in age group i and quintile q by $(1 - riskreduction_i)$ and then by 0.5.

$$frmrdhs_{iq} = fmrsmkr_{iq} * \\ (1 - riskreduction_i) * 0.5$$

      ii. For each quintile q, sum in every age group i the smokers prevented.

$$totfrmrdhs_q = \sum_{i=1}^{21} frmrdhs_{iq}$$

6.3.3 Total deaths averted

Smokers' deaths averted are calculated as the number of deaths averted from smokers who remain smoking and the deaths averted from smokers who quit due to the tax imposition. The first part is calculated as the difference between the expected deaths of total smokers in baseline and the expected deaths of total post-smokers. This result plus the number of deaths averted in smokers who quit results in the total smokers' deaths averted.

1. Method

   (a) BMJ/GTEC-Sub. Total deaths averted ($dthsaverted$) are the expected deaths of baseline smokers (who remain smoking) in q minus effective deaths of post smokers plus deaths averted among smokers who quit thanks to the intervention.

$$dthsaverted_q = (tot\_blsmkr_q * 0.5) \\ -(tot\_postsmkr_q * 0.5 + totfrmrdhs_q)$$

6.3.4 Deaths before tax

The total expected deaths before the tax increase is equal to the number of smokers in the baseline scenario times the expected deaths in this smokers that can be attributed to smoking.

1. Method

(a) BMJ/GTEC-Sub. Deaths before tax ($dths\_pre$) are the expected deaths due to the unchanged epidemic (people who remain smoking despite the intervention).

$$dths\_pre_q = tot\_blsmkr_q * 0.5$$

6.3.5 Deaths after tax

Deaths after the tax increase are calculated among the post-tax smokers. The proportion from these deaths attributable to tobacco consumption still remains being half (50%).

1. Method

   (a) BMJ/GTEC-Sub. Deaths after intervention ($dths\_post$): 50% of current smokers plus deaths averted among former smokers.

$$dths\_post_q = (totpostsmkr_q * 0.5) + totfrmrdhs_q$$

6.3.6 Share to total deaths

The share of total deaths due to the diseases modelled is calculated from colombian data sources. This is used to calculate the proportion from all tobacco attributable deaths which would occur due to each of the diseases included.

1. Source

   (a) BJM. Share to total deaths 2017 for copd, stroke, heart attack and lung cancer (tobacco attributable diseases) were obtained from the Institute for Health Metrics and Evaluation of the University of Washington. Shares used as input parameters were: 19% for copd, 22% for stroke, 52% for heart disease and 7% for lung cancer.

   (b) GTEC-Sub. Most recent share of deaths of tobacco attributable diseases (copd, stroke, heart attack and lung cancer) were obtained from Vital Statistics 2017 of National Department of Statistics [22]. The findings are summed up in Table 4 included in the Appendix.

2. Method

   (a) GTEC-Sub. Once deaths for every disease is obtained, single share to total deaths is computed.

$$share_d = \frac{deaths_d}{\sum deaths_d}$$

   Where $d = 1, 2, 3, 4$ represents copd, stroke, heart attack and lung cancer, respectively.

6.3.7 Disease's deaths averted

As mentioned before, using the shares from total deaths of each disease, the number of deaths averted that would occur due to each disease are computed.

1. Method

   (a) BJM/GTEC-Sub. Using total deaths averted and its corresponding disease share, deaths averted per cause of disease in each of the five income quintile are calculated.

   $$deaths_{qd} = dthsaverted_q * share_d$$

   Doing so, it is possible to determine how many of the total deaths averted were saved due to copd, stroke, heart attack and lung cancer in every quintile.

6.4 Healthcare

The healthcare module focuses in health services utilization and costs related to these utilizations in each of the four major diseases included in the model (COPD, stroke, heart disease and lung cancer). Copays, out of pocket expenditures and averted costs are also obtained.

6.4.1 Probability of seeking care

The first step into calculating healthcare expenditure corresponds to the health utilization of health services. To determine such use in the system, first individuals must reach the system (this involves choosing to use the system as well as having the means to reach it). Thus, the probability of seeking care becomes the first step in costs' calculations.

1. Source

   (a) BJM. Among the inputs used to develop the model listed in the supplement, the probability of seeking care due to each disease (u) does not have a singular source for Colombia inside the references. Apparently, authors used the same numbers of Turkey, they are 70% for all diseases [23].

   (b) GTEC-Sub. We use the same probabilities of seeking care as well as BMJ paper did in the Stata code for Colombia, it is 70% for all diseases.

6.4.2 Health utilization

Health utilization is related to the actual use of health services. This utilization of services differs among income distribution. Thus, health utilization is modeled using a gradient among the different income quintiles.

1. Source
   (a) BJM. Health utilization ($grad\_u$) by income quintiles were obtained from the working paper in which the impact of smoking on health systems in Latin America is studied [24]. Number used were as follows: 1 for q1, 1.1 for q2, 1 for q3, 1.1 for q4 and 1.2 for q5. However, it is not clear how and where BMJ paper authors exactly extracted the parameter from the cited document, a exhausted review of this source on our own allows us to confirm it.
   (b) GTEC-Sub. Due to the lack of clarity in inputs used, colombian ECEA team decided to make their own estimations about the gradient of utilization within income quintiles. To do this, we used ECV2018. Estimates by income quintile were: 0.9565 for q1, 0.9727 for q2, 0.9931 for q3, 1.0172 for q4 and 1.0603 for q5.
2. Method
   (a) GTEC-Sub. From Health chapter, health care utilization gradient was calculated per income quintile as follows:
      i. Health system utilization rate ($grad\_u$) was computed as the relative-to-mean proportion of people detected with a chronic disease (cardiovascular diseases, asthma, chronic bronchitis, gastritis, lupus, cancer, gout, leukemia, diabetes) who receive or received treatment to total people who were detected with such diseases. To point out, due to the specification of the question, it is not possible to discriminate by a particular disease as the ones studied in this document, so the result is a global utilization.
      $$util\_rate_q = \frac{detected\_treated_q}{detected_q}$$
      ii. Next, the gradient is found by quintiles as the relative-to-mean rate of the singular utilization rates.
      $$grad\_u_q = \frac{util\_rate_q}{\frac{\sum util\_rate_q}{5}}$$
      Calculation results are presented in Table 5 included in Appendix.

6.4.3 Treatment costs

Treatment costs are calculated for each of the diseases included. Costs are taken from different studies which have calculated them. Also a literature review was done searching for other possible sources for the costs used.

1. Source

   (a) BMJ. Annual treatment costs ($cost_d$) for the diseases considered (copd, stroke, heart attack and lung cancer) came out of previous economic research carried out for Latin America countries and their health systems. Direct costs were estimated as the sum of all caused expenditures necessary to attend the event. Costs in current USD of 2015 included in the model were: $361 for copd, $2.738 for stroke, $4.830 for heart attack and $13.223 for lung cancer [24].

   (b) GTEC-Sub. Literature was reviewed to account for disease costs. The disease name (lung cancer, COPD, heart disease and stroke) between quotation marks, followed by the keyword Colombia and costs (between quotation marks) was used as the search algorithm.

   Results obtained for each disease were reviewed excluding those results which did not have information on Colombia or no treatment costs. For COPD, treatment costs calculations were obtained for the year 2007 [25]. However, costs used in the BMJ taken are more recent, thus this estimations were chosen [24]. Regarding lung cancer, results obtained were mostly of incidence and prevalence estimates but lacked treatment costs. An estimation presented by the intelligence unit of The Economist was found using 2016 data [26]. Nonetheless, results presented turned out to be much higher when compared to those presented in Pichon-Riviere et al. Moreover, the study was sponsored by Roche laboratory and estimations were made using interpolation from Brazil data. Thus, for lung cancer estimates from Pichon-Riviere et al. were chosen as well.

   Treatment costs for stroke were searched for using the same algorithm. Results obtained gave three possible sources. Castañeda-Cardona et al. presents an estimated treatment cost following a cohort from december 2010 up to march 2013 [27]. Pichon-Riviere et al. gives an estimate in 2015 dollars using data from Argentina, Chile and Mexico [24]. Finally, an estimate for stroke between 2010 and 2012 is given by Camacho et al. [28]. Among these results, Pichon-Riviere et al. uses the most recent data and turns out as the most conservative estimate among these.

   Finally, estimates for heart disease are found in Camacho et al. as well [28]. The results they present turn out to be similar to those presented in Pichon-Riviere et al. However, the latter results use more recent data and are more conservative.

2. Method

(a) BMJ. Direct costs were estimated as the sum of all caused expenditures necessary to attend the event. Expressed in current USD of 2015. In calculations specifically for Colombia, in the Stata code authors used a ppp conversion factor of 1292 to allow for an international comparison and a exchange rate of COP $2.916. Neither the source of these numbers nor its period nor how they were calculated or what they represent (i.e. which one of all ppp conversion factors or if the exchange rate is a simple average or moving average) are specified within the document.

(b) GTEC-Sub. Taking this costs as BMJ paper did, we apply average exchange rate for 2018 (COP $2.955), according to official numbers of Banco de la República (Central Bank in colombia) to convert estimates into colombian currency. . Additionally, for a comparison in international dollars, 2018 World Bank ppp conversion factor of private consumption was used (1412).

6.4.4 Insurance coverage rate

Health utilization, and thus costs, depend on having health insurance and coverage of treatments. Insurance coverage rates are used as parameters towards health utilization and costs.

1. Source.

   (a) BMJ. According to a research about catastrophic expenditures in Colombia, in 2011 9.1% of the population was not insured neither in contributive nor subsidized health regime. In other words, left 90.9% was insured [29]. For this motive, this number was taken to the model as the insurance in health either for rural (ins_rur) and urban (ins_urb) areas.

   (b) GTEC-Sub. Insurance coverage rate for 2018 in Colombia was 94.44% [30], which was assumed either for rural (ins_rur) and urban (ins_urb) areas.

6.4.5 Financial support

Financial support as the proportion from costs which the government or institutions support population in health services used. For Colombia, due to its health system, this parameter takes a value of 1 (100%).

1. Source.

   (a) BMJ. Financial support in rural ($reim\_rur$) and urban ($reim\_urb$) was taken from the same paper as for insurance coverage rate [29]. The input value was 100%.

   (b) GTEC-Sub. We used the same number of financial support of 100% as BMJ did in inputs parameters for Colombia.

6.4.6 Copay

Copays are partial payments from treatment costs for which individuals are responsible. The amount for which individuals are responsible depends on health coverage and financial support given to individuals.

1. Method

   (a) BMJ/GTEC-Sub. Based on coverage rate and financial support of health system in Colombia, the copay rate ($copay$) is computed as follows:

$$
\begin{aligned}
copay = frac\_rur * (ins\_rur * (1 - reim\_rur) \\
+(1 - ins\_rur) * 1) + (1 - frac\_rur) \\
(ins\_urb * (1 - reim\_urb) + (1 - ins\_urb) * 1)
\end{aligned}
$$

   It means the amount of money that the person effectively pay for their treatment.

6.4.7 Expenditures paid out-of-pocket

Expenditures padi out-of-pocket (oop) for the colombian case correspond to the proportion of disease costs due to copays or to the proportion of health services not covered by the health system.

1. Method

   (a) BMJ/GTEC-Sub. Out-of-pocket expenditures (exp) paid individually for every disease can be calculated as the multiplication of copay rate and costs of treatment.

$$exp_d = copay * cost_d$$

6.4.8 Health care utilization by disease

Healthcare utilization is estimated as a function of the probability of using health services for each of the four diseases, and a gradient of utilization due to household income quintile.

1. Method

   (a) BMJ/GTEC-Sub. Chances of seeking healthcare and healthcare utilization after being detected by disease ($careuti$). Calculations are made for all diseases in 5 quintiles.

$$careuti_{qd} = grad\_u_q * share_d$$

6.4.9 Treatment costs averted without copay

Treatment costs averted are calculated as the costs not incurred in from people whose deaths were averted due to the tax imposition. Excluding out-of-pocket (i.e. copays) from the total expenditure results in the expenditure averted from treatment costs to the health system.

1. Method

   (a) BMJ/GTEC-Sub. Actual treatment costs averted without insurance ($treatcost$) are calculated for every disease and income quintile.

$$treatcost_{qd} = deaths_{qd} * cost_d * careuti_{qd}$$

6.4.10 Treatment costs averted including copay

Total costs are the sum of treatment costs including out-of-pocket or copays, in the colombian case. It is the total expenditure over the health system and the individuals as a whole.

1. Method

   (a) BMJ/GTEC-Sub. Treatment costs averted with copay ($avertedexp$) for all diseases and quintiles.

$$avertedexp_{qd} = deaths_{qd} * exp_d * careuti_{qd}$$

6.4.11 Out-of-pocket treatment cost despite copay

Treatment costs relying on the individual, or out-of-pocket costs, represent the costs which are paid directly by the individual through copays. This are calculated as the difference between total costs and treatment costs excluding copays.

1. Method

   (a) BMJ/GTEC-Sub. Out-of-pocket treatment costs in spite of copay (oop) for all diseases and income quintiles.

$$oop_{qd} = treatcost_{qd} - avertedexp_{qd}$$

6.5 Tax revenue

Tax revenue module, as it name suggests, is concerned with all variables necessary to calculate the total revenue from the new tax. Thus, this module contains variables regarding the quantity of cigarretes (in individual sticks which is transformed into number of packs), tax revenue from adults and young population, and the revenue obtained from the pre-tax scenario and post-tax scenario due to changes in consumption.

6.5.1 Consumption intensity

In order to be able to calculate the revenue generated due to the pre-tax and post-tax scenarios, the number of packs (which is the taxed good) must be calculated. However, tobacco consumption is obtained in number of cigarette sticks rather than cigarette packs. The number of (daily) cigarette sticks is transformed into number of packs by dividing the latter into 20. This number is used as in the calculation of tax revenue.

1. Source

   (a) BMJ. Number of cigarettes consumed per income quintile (cig) was obtained from ECSP 2013 [17]. Daily consumption was: 6 for q1, 8 for q2, 8 for q3, 8 for q4 and 10 for q5.

   (b) GTEC-Sub. Three different sources were consulted to build cigarette consumption intensity: DeicsCol 2017, ECV 2018 and ECSP 2013. Results are summed up in table 6 included in Appendix. For all sources, median was the statistic measure.

2. Method

   (a) BMJ. BMJ document used education lavel a proxy for income and they do not specify how they obtained the numbers about consumption or the methodology used.

   (b) GTEC-Sub. Owing to the itself design of the question asked in ECSP2013, it is not appropriate to capture the desired information about consumption. This difficulty is reflected in the answers in which all people surveyed reported just one cigarette smoked. Clearly this is not a faithful information, therefore, this source was dismissed.

   In DeicsCol Survey 2017, weekly consumption was computed assuming education as a proxy for income. To do so, highest educational level reported by the person is transformed into number of years studied. Assignment was as follows: 0 years for non-studies or pre-school, 5 years for primary, 9 for middle school, 11 for high school, 12 for technical or technological studies without degree, 13 for technical or technological studies with degree or college studies without degree, 16 for complete college studies, 17 for postgraduate studies without degree and 18 for postgraduate studies with degree. Next, national quintiles were computed. On the other hand, given that the geographical scope of the survey is only 5 principal cities in different regions of the country, information and calculations for these municipalities were assumed for the hole region. So, for instance, data of Medellín counts for Antioquia-Central region, Cali for Valle del Cauca, Cúcuta for Eastern region, Cartagena for Carribean region and Bogotá for Bogotá.

   In ECV 2018, income quintiles were defined from *percapita* variable. Then, number of cigarettes smoked per day were obtained (cig). Thanks to its geographical scope and its similarity in number of cigarettes smoked within quintiles to DeicsCol2017, ECV2018 was chosen as the final source for consumption intensity.

6.5.2 Price per pack

The next component necessary for tax revenue calculation is price (per pack). Once price is known, tax change can be calculated over the initial price and, using the income quintiles elasticities, changes in consumption can be calculated and tax revenue can be obtained as the new amount of daily packs and price.

1. Source
   (a) BMJ. BMJ paper authors used a price per pack ($price\_pack$) of 20 units COP $2.726. This was obtained as the median of all the 56 prices DANE certificated for the year 2016 for excise tax purposes [].
   (b) GTEC-Sub. The source to get prices was DeicsCol Survey 2017. Price per stick ($price\_stick$) was found first, COP $330, and then the correspondent price of 20 units pack, COP $6.600.

2. Method
   (a) GTEC-Sub. In order to obtain the price per pack of 20 units, prior calculations were necessary and are listed below:
      i. Using the t10 variable (number of smoked cigarettes weekly) and the population (expressed as the individual expansion factor) from DeicsCol Survey 2017, we calculated the total number of cigarettes smoked in the population.

$$weekly\_cig = cig * population$$

      ii. Taking stick prices already pre-computed in the survey for each observation and considering all distribution channels (big surfaces, neighborhood stores, liquor stores, street sellers, duty free, bars or restaurants, gas stations, San Andresito), we found a single stick “weighted” price as the multiplication of prices by total cigarettes divided by total cigarettes smoked.

$$price\_stick = \frac{\sum price\_stick * weekly\_cig}{\sum weekly\_cig}$$

      The reason to compute a weighted price is to take into account the differences in market size among cities included in questionnaire.

6.5.3 Baseline excise tax

Baseline excise tax is the tax in the pre-tax scenario. For Colombia this tax level is given by law and is equal to COP $2.100 for a 20 cigarette sticks pack.

1. Source
   (a) BMJ. BMJ paper authors used 49.5% as the share of tax to retail price to estimate baseline tax (bl_tax). The source for this number is not clear, though.
   (b) GTEC-Sub. According to article 347 of Law 1819/16, in which a Tax Reform is introduced in Colombia, the excise tax for cigarettes (bl_tax) for 2018 was established in COP $2.100 for a pack of 20 units or proportionally to its content [31].
2. Method
   (a) BMJ. In order to calculate the baseline tax, BMJ paper authors used the share of tax to retail price of 49.5%. Multiplying this share by the retail price, tax is found. The source for this number is not clear, though.

$$bl\_tax = tax\_share * retailprice$$

   (b) GTEC-Sub. We took the excise tax of COP $2.100 as established in Tax Reform 2016.

6.5.4 Post excise tax

Post excise tax is the new tax which is being modelled. In the colombian case, a scenario of increasing tax per 20 unit cigarette pack from COP $2.100 to COP $6.000.

1. Method
   (a) BMJ. Based on retail price adjusted for parity purchase power, baseline tax and the increase of 50% of price, post tax ($posttax$) is computed.

$$posttax = retailprice^{ppp} * pr\_inc + bl\_tax$$

   (b) GTEC-Sub. The main objective of this model is to assess the impact of an excise tax increase to the target tax ($target\_tax$) of COP $6.000. For this reason, changes were made in the formulation to achieve this goal. Baseline tax ($2.100) is multiplied by the needed times to increase it to reach $6.000 level (2.857 times).

$$posttax = bl\_tax * 2.857$$

6.5.5 Adult post smoker for tax revenue

To obtain the total revenue at a national and subnational level, the tax generated for each quintile must be aggregated by the number of smokers after the new tax. Total number of adult smokers are calculated separately from young smokers.

1. Method

   (a) BMJ/GTEC-Sub. Adult post smoker for tax revenue ($adult\_taxrev\_totpostsmkr$) are estimated from post smokers in every age group and income quintile (remain smoking after the intervention). Adult smokers are selected ($i > 3$) and summed for every income quintile q.

$$adult\_taxrev\_totpostsmkr_q = \sum_{i=4}^{21} postsmkr_{iq}$$

6.5.6 Young post smokers for tax revenue

Similarly to adult smokers, the total number of young smokers after the tax imposition are calculated.

1. Method

   (a) BMJ/GTEC-Sub. Young post smokers for tax revenue ($young\_taxrev\_totpostsmkr$) are estimated from post smokers in every age group and income quintile (remain smoking after the intervention). Young smokers are selected ($i = 5$) and summed for every income quintile q.

$$young\_taxrev\_totpostsmkr_q = \sum_{i=5}^{5} postsmkr_{5q}$$

6.5.7 Tax revenue prior intervention

Tax revenue prior to the tax change are calculated as the total number of daily cigarette packs in each quintile times 365 days to obtain the yearly revenue per smoker, and multiplying this result by the total number of baseline smokers and baseline tax.

1. Method

   (a) BMJ-GTEC-Sub. Depending on baseline adult smokers, daily consumption intensity and baseline excise tax, annual fiscal revenues before the intervention ($pretax$) are estimated per income quintile.

$$pretax_q = cig_q * (\frac{365}{20}) * bl\_tax * adult\_tot\_blsmkr_q$$

So as to guarantee consistency between estimates and official data, these numbers were validated with information about tax collection of every department in 2018, data consigned in Formato Único Territorial (FUT). Given the size of this form, only T.I.A. 1.18.1 and T.I.A. 1.18.4 subaccounts were considered for the calculations, as they contain information about the specific component tax of cigarettes. Table 10 in Appendix contains specific component tax collection reported by every department of the country.

Nevertheless, during this validation exercise, a very particular situation of tax collection in Bogota, described in detail in Appendix section, led us to introduce additional calculations in this module which were not included in the original ECEA model (also detailed in Appendix). The aim of this procedure was to get to meet ECEA estimates and Bogota's official information.

New calculations involved variables such as regional tax collection, 2017 and 2018 specific component tax rate and the interaction parameter: sales ratio of the first semester to the second (see Appendix to a more precise explanation). Although this was a particular situation only in Bogota, we decided to apply the methodology to all regions.

6.5.8 Tax revenue after intervention

Similarly, the pre-tax calculation is performed this time using the post-tax number of smokers as well as the modelled tax.

1. Method

   (a) BMJ-GTEC-Sub. Based on adult and young post smokers, daily consumption intensity and post excise tax, annual fiscal revenues after the intervention ($postax$) are estimated per income quintile.

$$\begin{aligned} postax_q = cig_q * (\frac{365}{20}) * posttax * \\ (adult\_taxrev\_totpostsmkr_q * cons_q + \\ young\_taxrev\_totpostsmkr_q * consyouth_q) \end{aligned}$$

6.5.9 Change in tax revenue

Finally, changes in tax revenue are calculated as the difference between the post-tax revenue and the pre-tax revenue.

1. Method

   (a) BMJ-GTEC-Sub. Given the pre and post tax revenues already estimated, taxes added ($taxadd$), thanks to the policy, is found per quintiles.

$$taxadded_q = postax_q - pretax_q$$

6.6 Poverty

In this module, relevant calculations are computed to describe the effect of undergoing to medical attendance owing to contracting any of the tobacco attributable diseases. Share of people falling into poverty or incurring in catastrophic expenditures as a result of these diseases jointly with the share of people averted of being part of these indicators are calculated using either simulated (BMJ) or observed (GTEC-Sub) personal income. Every procedure is specified next.

6.6.1 Income quintiles

Quintile boundaries are defined to classify population, useful to develop the rest of the module.

1. Source

   (a) BMJ. Due to the lack in information, authors simulated the annual personal income through a gamma distribution in which they used the average household income percapita and Gini index as inputs parameters to compute a general income simulation, as economic literature suggest [32]. COP $3,972,629 average income was obtained from DANE (National Department of Statistics, in english) and Gini index from World Bank Indicators. The number used in calculations for Colombia and consigned in Stata inputs code was 0.5. Its year is not specified.

   (b) GTEC-Sub. Contrary to BMJ's situation, observed monthly income was available for Colombia in two different sources: ECV2018 and Gran Encuesta Integrada de Hogares - GEIH2018 (Grand Integrated Household Survey). ECV2018 is the unique survey in charged of DANE that collects information about tobacco consumption and GIEH is the official and specially designed to measure labor market performance and monetary poverty as stipulated by national commandment [33]. Prevalence and consumption can be computed simultaneously in ECV2018, but its calculations in terms of income are not fully confident, because it is not its main objective. On the other hand, GEIH2018 contains very precise information about personal income but has no information about tobacco. Therefore, here arises the necessity to choose the most suitable and precise information to compute poverty module. To comply this, two comparisons are introduced: income and population distributions. In table 2 included in Appendix, the income distributions for both surveys are presented. Although in high percentiles they have considerable differences, in the middle ones they are quite similar. Table 3 contains the population percentage by income quintiles and age groups. Once again, both surveys depict close numbers across the groups.

   In conclusion, given the similarity in population distribution, secondly, that by definition income quintiles are comparable ordinarlly between both surveys (statistically the same amount of data is under q_i quintile) and that GEIH is particularly designed for labor market measurements, this last data source was elected to get income information from.

2. Method

   (a) BMJ. Once average income was ppp-adjusted, simulation is generated. For more mathematical details see supplementary appendix and in cited technical note. Once the incomes were settled, upper limits of each quintile are calculated.

   (b) GTEC-Sub. GEIH2018 collects individual's income information in 23 main cities, urban and rural areas to calculate poverty statistics. To do so, GEIH2018 constructs personal monthly income per unit expenditure as follows. First, any kind of individual monthly income is obtained: salaries, earnings, income in kind, pensions, government aids, etc. Then, household income is computed aggregating all members' income (excluding home maids, home maid's children, pensioners or home workers). Finally, this amount is divided by the number of people considering in the previous step. Usually, this calculated income is the same as the household income as a whole.

   Furthermore, observations reporting "suspecious" zero values in income were undergone to an special methodology of imputation whose final goal is to assign the real value of income. So, numbers reported in this survey and used in this section have been previously processed by DANE. If zero values appear in income list, is because in fact this is the real one, the person has null income. For methodological details see the complementary material of GEIH2018 [34]. Ultimately, using this personal income, quintiles are calculated jointly with its corresponding upper limits.

6.6.2 Statistics of income within quintiles

Descriptive statistics are calculated for each of the five income quintiles. In BMJ they are used to perform a second simulation but in GTEC-Sub they are just for informative purposes.

1. Method

   (a) BMJ. For every quintile q, mean (μ) and standard deviation (δ) of gamma-simulated income are found. Next, they are computed together to obtain alpha and beta paramenters.

$$\alpha_q = \frac{\mu_q^2}{\delta_q^2}$$
$$\beta_q = \frac{\delta_q}{\mu_q}$$

   These are used as inputs to further apply a second simulation of gamma distribution: a quintile-specific annual income. In other words, 5 different simulations were made, one for each quintile.

(b) GTEC-Sub. The same statistics as in BMJ are calculated but now using observed monthly personal income.

6.6.3 Poverty line

Including the poverty line in calculations allows to account for an official threshold which serves as the key reference to conclude upon the financial impact of suffering from a tobacco attributable disease in Colombia.

1. Source

   (a) BMJ. Although BMJ paper authors mentioned in the document that they used the World Bank's ppp (parity purchase power) adjusted poverty line ($povline$) of USD $1.90 a day, in the Stata code they used a poverty line of USD $1.96, not clear why [35]. Here a concern arises when calculating the annual poverty line. So as to make all monetary units meet, daily poverty line is transformed to get an annual standard, it is 365 times the poverty line. Consequently, depending on the value used, the result will differ. If it is $1.90, yearly line is USD $693.5 and $715.4 if it is $1.96.

   (b) GTEC-Sub. Colombian poverty line ($povline$) used in the model was the total national one of 2018 of COP $257,433, as reported by National Department of Statistics [36]. This number reflects the minimal amount of personal money needed to purchase a basic basket of alimentary goods that complies with the nutritional recommendations of FAO/WHO/UNO. For more methodological details see the Misión para el Empalme de las Series de Empleo, Pobreza y Desigualdad -MESEP appendix [34].

6.6.4 Impoverished due to disease

Share of people falling into poverty owing to disease treatment costs is calculated by income quintile. It is, what percent of population go below the poverty line thanks to inccuring in treatment of one of the diseases considered: copd, stroke, heart attack, lung cancer.

1. Method.

   (a) BMJ. Impoverished due to disease ($impoverished$) procedure is repeated for the 5 quintile-specific simulations and is as follows:

      i. Annual cost-specific treatment of every tobacco attributable disease is substracted from annual simulated income in order to find available income after undergoing to medical treatment.

ii. If initial income was above the poverty line and latter was below it, the person is considered as impoverished due to the specific disease and marked 1. Otherwise, if the person was at the beginning poor or if the treatment does not set them poor, the person is marked 0.

iii. Finally, the average of this dichotomic variable is calculated by income quintile. This gives the the percentage of people, in each income group, falling into poverty thanks to the disease they acquired for smoking.

(b) GTEC-Sub. Impoverished due to disease ($impoverished$) was the same procedure applied before but now using observed income. Moreover, some calculations were modified. Changes are listed next.

i. Given that the income used is monthly, to substract annual treatment cost we multiply income by 12. The result was divided by 12 to obtained monthly income available to make it comparable to the poverty line in Colombia.

ii. Left steps remain the same as in BMJ.

6.6.5 Catastrophic out-of-pocket expenditures

Share of people inccuring in catastrophic expenditures is calculated by income quintiles and for every of the four diseases considered. An expenditure is considered as catastrophic when it comprises more than 10% of the individual's yearly income [37].

1. Method

(a) BMJ. People incurring in catastrophic expenditures due to costs treatment ($cata$) were found for every of the four diseases considered.

i. Share of treatment cost to yearly income was calculated as the division of income to cost. If costs exceed 10% of income, the person is marked 1, otherwise, it is marked 0. The procedure is done for all diseases.

ii. The average of this dummy variable is calculated by income quintile. This returns the percentage of people who incur in catastrophic expenditure owing to medical treatment.

(b) GTEC-Sub. To generate catastrophic expenditures ($cata$), costs was divided by 12 times monthly income, to compute equal units. Observations reporting zero in income were automatically assigned the value of 0.11 (11%), meaning that disease costs comprise more than 10% of their income. Consequently, they are marked 1 and counted as people inccuring in catastrophic expenditures. The rest of the procedure remains the same as in BMJ.

6.6.6 Non-poor people

Non-poor people is calculated by income quintile comparing estimated poverty line with simulated income (BMJ) and official poverty line with observed income (GTEC-Sub). The decision criterion is: if personal income is below the poverty line, the person in considered poor, otherwise, they are not.

1. Method

   (a) BMJ/GTEC-Sub. Applying the criterion, poor were marked 0 and non-poor ($nonpoor$) 1. Next, estimating the average of this variable, poor people share is found per quintiles. The interpretation: the closer to 1, the lower the poverty rate and, conversely, the nearer to 0, the higher it is.

6.6.7 Poor people averted due to diseases

Number of poor people averted due to each disease is found by income quintile.

1. Method

   (a) BMJ/GTEC-Sub. Share of people averted from impoverishing due to disease ($avertedimp$) is multiplied by the number of particular-disease deaths and then by health care utilization.

$$avertedimp_{qd} = impoverished_{qd} * deaths_{qd} * careuti_{qd}$$

6.6.8 People averting catastrophic expenditures due to diseases

Number of people averting catastrophic expenditures due to each disease is found by income quintile.

1. Method

   (a) BMJ/GTEC-Sub. Share of people prevented from inccuring in catastrophic expenditures due to disease ($avertedcata$) is multiplied by the number of particular-disease deaths and then by health care utilization.

$$avertedcata_{qd} = cata_{qd} * deaths_{qd} * careuti_{qd}$$

6.7 Macro-consistency

In the end, the model exactly specified as described along this report, did not have a macro-consistency between subnational to national level (i.e. the sum of subnational equals the national). To identify the cause, two different scenarios were proposed, in which a validation process was done, requiring an exhausted revision across every input used in the model and its respective outcomes. During the checking exercise and given the inconsistencies found, some changes were necessary to reach our goal. Intermediate steps, jointly with the decisions made in each one, are listed next.

1. The first, the baseline one, considered all the input's source (prevalence and population for every age group) in ECV2018, involving a change in age group population source. At the very beginning, age group population was obtained from DANE projections for year 2018 based on numbers of census 2005. This became an issue when computed with the age-group prevalence (ECV2018) because total population did not coincide in both sources (49 834 240 people in projections and 49 987 281 in ECV2018). So, this would lead to a mismatch in subnational to national closure. So, source-harmonizing would allow us to start from a 'theoretical' exact match.

   Surprisingly, perfect macro-consistency did not occur in all the variables in the model, just in some of them (adult, young and total smokers). Nevertheless, it is worth saying that the relative mismatch is 0.0001% in all the variables related to treatment costs and less than 1% in the remaining. However, tax revenue variables (pre, post and added taxes) had a difference of about 12%.

2. Keeping previous modifications, second scenario consisted in adjusting young prevalence with ECSPPE2016, that is prevalence in 10-14 and 15-19 age groups. Doing so, any of the variables reported a perfect closure as before, treatment variables worsened but the mismatch never went up more than 1%, excepting the tax-related ones, which maintained the 12% difference. Here, the gap is explained due to the incorporation of an estimate obtained from a different source, with different population size, because it is only considering people ranging from 12 to 18 years.

While developing the validations, macro-consistency among the total number of smokers showed no macro-consistency. This was due to differences between the population distribution among the subnational level in DANE's population projections and population distribution in ECV. Although there is population macro-consistency when using population projections, once prevalences obtained from ECV are applied to such populations, macro-consistency among number of smokers is lost in all age-groups.

In summary, best adjustment (not even perfect) is reached with a model that underestimate young smokers (baseline scenario), tax-related variables never got a better fit and current model does not comply with macro-consistency. Consequently, to guarantee the macro-closure at the same time of preserving reasonable estimates in young population and taking into account that DANE surveys are built upon their own annual population projections, which means that ECV2018 implicitly contains DANE estimations, we no longer use population from projections and rather take it from ECV2018. At the national level we no longeer make independent calculations and it becomes the result of aggregating all the subnational variables.

6.8 Indicators

6.8.1 Tax revenues

So as to expose the fiscal gains of increasing the excise tax of cigarettes to COP $6.000 with policymakers, specially governors, two indicators of interest were developed.

1. **Departamental taxes added**. In order to disaggregate regional tax revenue after the tax hike, departmental to regional smokers' share was calculated. This metric allows to find single-department revenue while preserving the representativeness (market size) of each territory. Percentages are in table 9 included in Appendix.

   $$dpto\_sh_{jr} = \frac{dpto\_smkr_{jr}}{reg\_smkr_{r}}$$

   The result is the smoker share of department j in region r. Applying this to regional tax revenue, is possible to tell governors what the new level of fiscal collection will be when adopting the intervention.

2. **Tax revenue gap**. Using observed (official numbers from Formato Único Territorial) and pre-intervention tax revenue (output of ECEA model based on consumption intensity and applying $dpto\_sh_{jr}$ to disaggregate), tax revenue gap was calculated by department.

   $$dpto\_taxgap_{j} = obs\_taxrev_{j} - reg\_taxrev_{r} * dpto\_sh_{jr}$$

7 Subnational ECEA

7.1 Definition of regions

7.1.1 Criteria

Sub national regions were assumed the same way as the National Department of Statistics defined them in the documentation for the Quality of Life Survey implementation, which turns to be one of the main data sources for this study. The list is specified in 1 bellow:

Table 1: Regions' definition

Region	Departments
Bogotá	Bogotá D.C.
Antioquia-Central	Antioquia, Caldas, Quindío, Risaralda, Tolima, Huila, Caquetá
Caribe	Guajira, Cesar, Magdalena, Atlántico, Bolívar, Sucre, Córdoba
Valle	Valle del Cauca
Oriental	Norte de Santander, Santander, Boyacá, Cundinamarca, Meta
Otras	Chocó, Cauca, Nariño, Arauca, Casanare, Vichada, Guainía, Guaviare, Vaupés, Amazonas, Putumayo, San Andrés

7.2 Model

The aim of the model is to examine the the impact in health, poverty and additional tax revenues resulted from a rise in cigarette excise tax from current COP $2.100 to COP $6.100 in Colombia. Same model as the one in the BMJ paper

Assumptions

References

[1] INSP. *The Global Tobacco Economics Consortium. Strengthening the leadership of Mexico, Colombia and India to advance tobacco control policy*. Tech. rep. Instituto Nacional de Salud Publica (INSP), 2018.

[2] Verguet, Stéphane, Laxminarayan, Ramanan, and Jamison, Dean T. "Universal Public Finance of Tuberculosis Treatment in India: An Extended Cost-Effectiveness Analysis". In: *Health Economics* 24.3 (Feb. 2014), pp. 318–332.

[3] Verguet, Stéphane, Kim, Jane J, and Jamison, Dean T. "Extended Cost-Effectiveness Analysis for Health Policy Assessment: A Tutorial". In: *PharmacoEconomics* 34.9 (2016), pp. 913–923.

[4] James, Erin et al. "The distributional consequences of increasing tobacco taxes on Colombia's health and finances. An extended cost-effectiveness analysis". 2017.

[5] James, Erin et al. "Distributional health and financial benefits of increased tobacco taxes in Colombia: results from a modelling study." In: *Tobacco control* (2018), tobaccocontrol–2018–054378.

[6] Chaudhuri, Arijit. *Modern Survey Sampling*. CRC Press, 2014. ISBN: 978-1-4665-7261-4.

[7] Levy, Paul and Lemeshow, Stanley. *Sampling of Populations. Methods and Applications*. 4th Ed. Wiley, 2008.

[8] Global Tobacco Economics Consortium. "The health, poverty, and financial consequences of a cigarette price increase among 500 million male smokers in 13 middle income countries: compartmental model study". In: *BMJ* 361 (Apr. 2018), k1162.

[9] IARC. *Vol. 14: Effectiveness of Tax and Price Policies for Tobacco Control*. Tech. rep. International Agency for Research on Cancer - IARC, 2011.

[10] NCI and WHO. *The Economics of Tobacco and Tobacco Control*. Tech. rep. NCI Tobacco Control Monograph Series 21. Department of Health and Human Services, National Institutes of Health, U.S National Cancer Institute (NCI), World Health Organization (WHO), 2016.

[11] Gallego, Juan Miguel, Rodriguez-Lesmes, Paul, and Otalvaro-Ramirez, Susana. "Price smoking participation elasticity in Colombia: Estimates by age and socio-economic level". 2019.

[12] Maldonado, Norman, Llorente, Blanca, and Deaza, Javier. "Impuestos y demanda de cigarrillos en Colombia". In: *Rev Panam Salud Publica* 40.4 (2016), pp. 229–36.

[13] MinSalud. *Impuestos al tabaco*. Tech. rep. Papeles en salud, edicion No. 1. Ministerio de Salud y Proteccion Social, Republica de Colombia, 2016.

[14] Araque, Alex. "Tabaco en Colombia: costos y ¿beneficios?" In: *Civilizar: Ciencias Sociales Y Humanas* 14.26 (2014), pp. 133–148.

[15] James, Erin et al. “The distributional consequences of increasing tobacco taxes on Colombia’s health and finances: An extended cost-effectiveness analysis”. 2017.

[16] James, Erin Kinsella et al. “Distributional health and financial benefits of increased tobacco taxes in Colombia: results from a modelling study”. In: *Tobacco Control* 28.4 (July 2019), 374 LP - 380.

[17] MinJusticia, MinSalud, and O.D.C. *Estudio Nacional de Consumo de Sustancias Psicoactivas en Colombia 2013. Informe Final*. Tech. rep. Bogotá: Ministerio de Justicia y del Derecho (MinJusticia), Ministerio de Salud y Protección Social (MinSalud) y Observatorio de Drogas de Colombia (O.D.C.), 2014.

[18] UN. “World Population Prospects: The 2015 Revision, Methodology of the United Nations Population Estimates and Projections”. 2015.

[19] Verguet, Stéphane et al. “Distributional benefits of tobacco tax and smoke-free workplaces in China: A modeling study”. eng. In: *Journal of global health* 7.2 (Dec. 2017), p. 20701.

[20] Doll, Richard et al. “Mortality in relation to smoking: 50 years’ observations on male British doctors”. In: *BMJ* 328.7455 (June 2004), p. 1519.

[21] Jha, Prabhat et al. “21st-Century Hazards of Smoking and Benefits of Cessation in the United States”. In: *New England Journal of Medicine* 368.4 (Jan. 2013), pp. 341–350.

[22] DANE. *Estadísticas vitales - EEVV. Cifras definitivas año 2017*. Tech. rep. Dirección de Censo y Demografía, Departamento Administrativo Nacional de Estadística (DANE), República de Colombia, 2018.

[23] OECD. *OECD Reviews of Health Care Quality: Turkey 2014*. 2014, p. 140.

[24] Pichon-Riviere, Andrés et al. “Economic impact of smoking on health systems in Latin America: A study of seven countries and its extrapolation to the regional level”. In: *Pan American Journal of Public Health* 40.4 (2016).

[25] Ciapponi, Agustin et al. “The epidemiology and burden of COPD in Latin America and the Caribbean: systematic review and meta-analysis”. In: *COPD: Journal of Chronic Obstructive Pulmonary Disease* 11.3 (2014), pp. 339–350.

[26] Kielstra, Paul. *Lung cancer in Latin America: Time to stop looking away*. Tech. rep. The Economist Intelligence Unit, 2018.

[27] Castañeda-Cardona, Camilo et al. “Análisis de costos de atención de infarto cerebral agudo con o sin fibrilación auricular”. In: *Acta Neurológica Colombiana* 30 (2014), pp. 78–82.

[28] Camacho, Sandra et al. “How much for a broken heart? Costs of cardiovascular disease in Colombia using a person-based approach”. In: *PLOS ONE* 13.12 (Dec. 2018), e0208513.

[29] Amaya-Lara, Jeannette Liliana. “Catastrophic expenditure due to out-of-pocket health payments and its determinants in Colombian households”. In: *International Journal for Equity in Health* 15.1 (Nov. 2016), p. 182.

[30] Minsalud. *Análisis de Situación de Salud (ASIS). Colombia, 2018*. Tech. rep. Dirección de Epidemiología y Demografía, Ministerio de Salud y Protección Social (Minsalud), República de Colombia, 2018.

[31] Congreso de la República. *Ley 1819 de 2016*. 2016.

[32] Kemp-Benedict, Eric. *Income distribution and poverty methods for using available data in global analysis*. Tech. rep. PoleStar Technical Notes No. 4, 2001.

[33] DNP. *Metodologías oficiales y arreglos institucionales para la medición de la pobreza en Colombia (Conpes Social 150)*. Tech. rep. Consejo Nacional de Política Económica y Social. República de Colombia, Departamento NAcional de Planeación (DNP), 2012.

[34] DNP and DANE. *Misión para el Empalme de las Series de Empleo, Pobreza y Desigualdad (mesep). Pobreza monetaria en Colombia: Nueva metodología y cifras 2002-2010. Resultados segunda fase de la Mesep*. Departamento Nacional de Planeación (DNP), Departamento Administrativo NAcional de Estadística (DANE), República de Colombia, 2012. ISBN: 978-958-624-095-6.

[35] World Bank. *Poverty and Shared Prosperity 2018: Piecing Together the Poverty Puzzle*. Washington, DC: World Bank, 2018. ISBN: 978-1-4648-1330-6.

[36] DANE. *Pobreza monetaria en Colombia*. Tech. rep. Boletín técnico COM-030-PD-001-r-004 V8. Departamento Administrativo Nacional de Estadística (DANE), República de Colombia, 2018.

[37] WHO. *The World Health Report. Health Systems Financing: The path to universal coverage*. World Health Organization (WHO), 2010. ISBN: 978 92 4 156402 1.

Appendices

A Methodology

A.1 Income quintiles

National income distribution of ECV2018 and GEIH2018 is presented in Table 2:

Table 2: Income distribution ECV vs GEIH

	ECV2018	**GEIH2018**
min	0	0
p25	209,938	231,154
p50	395,067	416,667
p75	775,000	754,550
p90	1,666,667	1,357,778
p95	2,824,000	2,040,984
p99	7,166,667	4,744,444
max	150,450,000	88,833,333
mean	801,366	679,391

All values in current COP.

Age distribution (life cycles) at national level was computed within each income quintile. Groups were defined as Ministry of Health does, as follows: early childhood (0-5 years), childhood (6-11 years), adolescence (12-17 years), adult (18-64 years) and elderly (65 years or more). Table 3 contains the results:

Table 3: Population distribution by life cycles

		Life cycle				
Q	**Source**	**0-5**	**6-11**	**12-17**	**18-64**	**>64**
1	ECV	14.2	15.7	13.5	49.6	7.1
	GEIH	14.6	16.4	14.2	48.7	6.1
2	ECV	13.7	14.9	13.5	51.9	6.1
	GEIH	12.1	12.6	12.3	56.4	6.6
3	ECV	11.9	11.9	11.8	57.9	6.4
	GEIH	9.6	10.1	10.1	62.6	7.6
4	ECV	8.9	9.3	9.5	64.4	8.0
	GEIH	6.8	7.7	7.7	68.2	9.6
5	ECV	5.7	6.1	5.9	70.3	12.0
	GEIH	4.4	4.9	4.9	72.2	13.6

Pop in i quintile and j life cycle / total pob in i quintile.

A.2 Prevalence

According to National Psychoactive Substances Survey 2013 for the age group 12-65 years, men had a higher prevalence than women, 18.84% and 7.4%, respectively, pointing out a gap of about 10pp. On the other hand, most recent Quality of Life Survey (2018) reports a prevalence of 13.8% for men and 4% for women, maintaining the 10pp gap.

A.3 Diseases' deaths

Table 4: Diseases' deaths 2017

Disease	Deaths	%
Copd	1,052	0.0437
Stroke	426	0.0177
Heart attack	21,714	0.9035
Lung cancer	840	0.0349
Total	24,032	

Source: Vital Statistics 2017, DANE.

A.4 Gradient of health utilization

National relative-to-mean gradient of health utilization is presented below. Table shows number of people responding affirmative in receiving treatment once they have been detected with a chronicle disease.

Table 5: Gradient of health utilization

Treatment	q1	q2	q3	q4	q5	Total
Yes	658,746	805,554	1,042,362	1,467,993	2,450,147	6,424,801
No	126,766	139,024	154,864	178,148	185,636	784,440
Total	785,512	944,578	1,197,226	1,646,142	2,635,784	7,209,241
Ut. rate	0.838	0.852	0.87	0.891	0.929	0.891
mean ut. rate	0.8766					
gradient ut.	0.956	0.972	0.993	1.017	1.060	

Source. ECV 2018, DANE.

A.5 Cigarette consumption

Table 6: Cigarettes consumption

Source	Income	p25	p50	p75
	1	28	49	70
	2	14	35	70
DeicsCol2017	3	14	28	49
	4	14	28	56
	5	14	35	56
	total	14	35	63
	1	3	5	10
	2	3	5	10
ECV2018	3	3	5	10
	4	3	5	10
	5	3	5	10
	total	3	5	10
	1	1	1	1
	2	1	1	1
ENCSPA2013	3	1	1	1
	4	1	1	1
	5	1	1	1
	Total	1	1	1

In DeicsCol2017, numbers reflect cigarettes consumed weekly. In ENCSPA2013, the question asked was: Approximately, how many cigarettes have you smoked daily in the last 30 days? Its design is not appropriate to capture information about consumption behaviour because, in the first place, reporting something done in the last 30 days is not easy to remember and given this uncertainty, the incentives to underreport increases.

A.6 Household characterization

Table 7: Household typology ECV2018

Typology	**Number**	**%**
Unipersonal	2,440,805	15.8
Men (M)	1,443,658	9.3
Women (W)	997,147	6.4
M-W	1,605,848	10.4
M-W 1 offspring	2,214,364	14.3
M-W 2 offspring	2,092,482	13.5
M-W 3 offspring	748,082	4.8
M-W 4 offspring	209,541	1.4
Women single parent	400,756	2.6
Households in survey	15,493,441	

The definition of women single parent was according to colombian Law 82/1993.

A.7 Share of smokers by department

Departmental to country smokers share is computed using two different sources: ECV18 and Encuesta de Consumo de Sustancias Psicoactivas 2013 - ECSP2013 (National Psychoactive Substances Survey). The reason was to make a data cross-validation and to evidence any possible changes in epidemic behaviour. Table 9 contains the departmental to regional smokers share, which was used to assign regional tax revenue to a particular department.

Table 8: Departmental to national smokers share

Department	Smokers >=18 ecv2018	%	Smokers >=18 encspa2013	%	Smokers <18 encspa2013	%
Antioquia	532,136	17.66	589,251	20.60	13,193	8.29
Atlántico	95,639	3.17	92,672	3.24	-	-
Bogotá	639,454	21.23	886,330	30.99	62,784	39.44
Bolívar	90,097	2.99	67,289	2.35	8,195	5.15
Boyacá	66,031	2.19	39,382	1.38	7,867	4.94
Caldas	84,224	2.80	61,083	2.14	1,660	1.04
Caquetá	15,721	0.52	19,202	0.67	1,669	1.05
Cauca	58,779	1.95	22,975	0.80	1,143	0.72
Cesar	43,560	1.45	20,057	0.70	-	-
Córdoba	70,022	2.32	35,354	1.24	-	-
Cundinamarca	174,889	5.81	146,053	5.11	8,565	5.38
Chocó	21,341	0.71	5,803	0.20	748	0.47
Huila	48,657	1.62	37,845	1.32	1,488	0.93
La Guajira	45,923	1.52	16,662	0.58	625	0.39
Magdalena	60,223	2.00	43,952	1.54	1,208	0.76
Meta	63,093	2.09	54,022	1.89	6,198	3.89
Nariño	103,612	3.44	42,867	1.50	5,327	3.35
Nte. Santander	71,585	2.38	81,690	2.86	2,022	1.27
Quindío	39,796	1.32	35,494	1.24	3,009	1.89
Risaralda	63,792	2.12	70,610	2.47	8,623	5.42
Santander	119,247	3.96	104,008	3.64	9,568	6.01
Sucre	49,716	1.65	18,133	0.63	2,474	1.55
Tolima	94,686	3.14	41,092	1.44	2,507	1.58
V. Cauca	303,800	10.08	289,056	10.11	8,467	5.32
Arauca	7,705	0.26	10,604	0.37	1,134	0.71
Casanare	19,159	0.64	10,201	0.36	-	-
Putumayo	6,867	0.23	4,200	0.15	-	-
San Andrés	1,924	0.06	4,406	0.15	-	-
Amazonas	3,791	0.13	3,050	0.11	-	-
Guainía	3,122	0.10	2,606	0.09	-	-
Guaviare	6,328	0.21	2,450	0.09	699	0.44
Vaupés	5,537	0.18	1,743	0.06	-	-
Vichada	2,264	0.08	156	0.01	-	-
TOTAL	**3,012,720**		**2,860,297**		**159,172**	

Table 9: Departmental to regional smokers share

Region	**Smokers >=18**	**%**
Antioquia-Central	879,013	
Antioquia	532,136	60.54
Caldas	84,224	9.58
Caquetá	15,721	1.79
Huila	48,657	5.54
Quindío	39,796	4.53
Risaralda	63,792	7.26
Tolima	94,686	10.77
Caribe	455,180	
Atlántico	95,639	21.01
Bolívar	90,097	19.79
Cesar	43,560	9.57
Córdoba	70,022	15.38
La Guajira	45,923	10.09
Magdalena	60,223	13.23
Sucre	49,716	10.92
Oriental	494,844	
Boyacá	66,031	13.34
Cundinamarca	174,889	35.34
Meta	63,093	12.75
Nte. Santander	71,585	14.47
Santander	119,247	24.10
Otras	240,429	
Cauca	58,779	24.45
Chocó	21,341	8.88
Nariño	103,612	43.09
Arauca	7,705	3.20
Casanare	19,159	7.97
Putumayo	6,867	2.86
San Andrés	1,924	0.80
Amazonas	3,791	1.58
Guainía	3,122	1.30
Guaviare	6,328	2.63
Vaupés	5,537	2.30
Vichada	2,264	0.94

A.8 Tax revenue

Observed tax revenue from the specific component in Bogota was COP $175.2 billion. In order to estimate the implicit number of packs, the revenue as divided by the specific tariff of COP $2,100. But this resulted in a very low number: 83 million of 20-unit packs. During validation with Bogota's tax authority, the staff informed of a relevant: during the first semester the distributors calculated the payments using the 2017 tariff (COP $1,400). To capture this situation and obtain a more realistic estimates of the implicit number of packs, we applied the following steps:

1. We included the parameter (m) that captures the assumption of sales ratio between the first semester to the second. For instance, if the parameter takes the value of one, it means that volume of sales is the same in the two periods. Values above 1 indicate that more than 50% of sales take place during the first semester, while values below 1 are consistent with the majority of sales occurring during the second semester.

2. Revenues result form the following equation:

$$total\_rev = 1,400 * Q_{sem1} + 2,100 * Q_{sem2}$$

   Where:

   $Q_{sem1} = mQ_{sem2}$.

3. The volume of packs for the second semester is:

$$Q_{sem2} = \frac{total\_rev}{1,400m + 2,100}$$

4. Using the information available for Bogota and assuming $m = 1$, $Q_{sem2} = 50$ million packs. Then, total number of packs for the whole year is 100 million.

5. To establish the difference between the packs that pay tax and the packs consumed by smokers, the number obtained in step 4 was compared to the number of packs estimated by smoker surveys (ECV2018 adjusted with prevalence for youths from ECSPPE2016).

Table 10: Specific component tax collection 2018

Deparment	**Million COP**
Amazonas	2,257
Antioquia	118,050
Arauca	1,439
Atlantico	24,020
Bogota	175,200
Bolivar	19,395
Boyaca	17,177
Caldas	18,111
Caqueta	7,300
Casanare	2,775
Cauca	5,385
Cesar	3,869.9
Choco	3,193
Cordoba	12,262
Cundinamarca	65,630
Guainia	1,089
Guaviare	1,249.6
Huila	19,401
La Guajira	2,608
Magdalena	5,060
Meta	20,990
Nariño	36,600
N. Santander	11,828
Putumayo	5,111
Quindío	18,480
Risaralda	30,270
San Andrés y Prov.	486.8
Santander	3,3160
Sucre	3,779
Tolima	23,070
V. del Cauca	98,560
Vaupes	396
Vichada	335.2
Total national	788,537.5