<u>Appendix 1.</u> Sampling strategy used for the *PLS* survey (or the *FTF-household survey*) based on the probability of smoking

The survey sample "target population" includes all census tracts with at least 50 private households covering the entire territory of the target cities. Neighboring census tracts with less than 50 private households, given the chance of not finding a sufficient number of smokers to complete the sample or to find littered packs in the streets, were grouped into a new primary sampling units (PSU) with at least 150 private households.

The sample design of the study consisted of two stages of selection for Rio de Janeiro and São Paulo and one stage for the other cities.

- First stage of selection:

In the first stage, 70 PSU were selected with probability proportional to the estimated number of smokers residing in each census tract (NSK).

As the Demographic Census does not provide information on the number of smokers living in the census tracts. this information was estimated as follows:

First, for each city, we used the individual data from the National Health Survey (*PNS*) survey conducted in 2013 to model the probability of smoking using a logistic regression model (see equation below):

$$p_i = \frac{exp(\beta_0 + \beta_1 SX_i + \beta_2 FE1_{i1} + \beta_3 FE2_i + \beta_4 FE3_i + \beta_5 FE4_i + \beta_6 FE5_i + \beta_7 RD_i)}{1 + exp(\beta_0 + \beta_1 SX_i + \beta_2 FE1_{i1} + \beta_3 FE2_i + \beta_4 FE3_i + \beta_5 FE4_i + \beta_6 FE5_i + \beta_7 RD_i)}$$

, where:

pi is the estimated probability of an individual i being a manufactured cigarette smoker;

SXi is the gender of individual i [male (0) and female (1)];

FE1i assumes value 1 if the age of individual i is between 25 and 34 years old and assumes value 0 for the other age;

FE2i assumes value 1 if the age of individual i is between 35 and 44 years old and assumes value 0 for the other age;

FE3i assumes value 1 if the age of individual i is between 45 and 54 years old and assumes value 0 for the other age;

FE4i assumes value 1 if the age of individual i is between 55 and 64 years old and assumes value 0 for the other age;

FE5i assumes value 1 if the age of individual i is equal to or greater than 65 years and assumes value 0 for other age;

FE1i. FE2i. FE3i. FE4i and FE5i. assume value 0 if the individual i is between 18 and 24 years old;

RDi is the average monthly income of the household owner of individual I, including those who also have no income;

β0. β1. β2. β3. β4. β5. β6 and *β7* are the intercept and coefficients of the adjusted model.

The final models for each city included all variables, regardless of the statistical significance of the coefficient.

Next, residents aged 18 and over from each census tract were distributed by sex (male and female) and age group (18-24; 25-34; 35-44; 45-54; 55-64 and 65+) based on the National Census conducted in 2010. For each combination of gender and age group (12 categories), the prevalence of smokers (*pj*) was estimated, given census tracts´ specific average monthly income of the household owner.

The estimated number of smokers in each census tract by sex and age groups was calculated by the product between the corresponding prevalence of smokers and the

number of people aged 18 and over residing in the census tract in the same sex-age group category. Finally, the expected number of smokers in each census tract calculated by summing the estimated number of smokers in each sex-age group category, as follows:

$$NSK_j = \sum_{k=1}^{12} p_{jk} * POP_{jk}$$

, where:

j -> represents the census tract;

k -> ranges from 1 to 12 and represents gender and age group;

NSK -> is the estimated number of smokers in sector j;

Pjk -> is the estimated prevalence for category k of sector j;

POPjk -> is the number of persons aged 10 years and over residing in census tract j and belonging to category k.

- Second stage of selection (only for face to face household survey (FTF-household)):

The *FTF-household* survey was only conducted in Rio de Janeiro and São Paulo cities. We used information on the gender-age distribution of the smoking population living in the same census tracts selected in the first stage of selection to determine the quota of smokers to be interviewed (n=11 per census tract).