



ITC France Survey Waves 1-3 (2006-2012) Technical Report

Prepared by C. Boudreau, M. Tait, & G. Li

Dept. of Statistics & Actuarial Science and Data Management Core (DMC) of ITC Project

University of Waterloo

Waterloo, Ontario, N2L 3G1 Canada

Updated August 2016

Suggested Citation

ITC Project (2016, August). *ITC France Waves 1 to 3 (2006-2012) Technical Report*. University of Waterloo, Waterloo, Ontario, Canada; French Institute for Health Promotion and Health Education (INPES), and French National Cancer Institute (INCa), Paris, France.

Sampling Design and Weight Construction for the International Tobacco Control (ITC) France Survey

Fourth Edition (Waves 1–3)

C. Boudreau^{1,2}, M. Tait^{2,3} & G. Li^{2,3}

This technical report details the sampling design and weight construction for waves 1–3 of the International Tobacco Control (ITC) France Survey. The ITC France Survey is a prospective longitudinal survey of a nationally representative random sample of adults smokers and non-smokers residing in continental France.

This technical report is organized as follows: section 1 describes the sampling design of the ITC France Survey, and section 2 details the construction of the sampling weights for wave 1 (section 2.2), wave 2 (section 2.3), and wave 3 (section 2.4).

1 Sampling Design

The ITC France Survey is a prospective longitudinal study, and its sampling design was chosen to yield representative random samples of adult smokers and non-smokers residing in that country. Respondents were first interviewed in Dec. 2006–Feb. 2007 (wave 1), with follow-up interviews in Sep.–Dec. 2008 (wave 2) and in Sep.–Dec. 2012 (wave 3). Those lost to follow-up at waves 2 and 3 were replenished by new randomly selected respondents. All interviews were conducted using computer assisted telephone interviews (CATI).

To qualify for the study, respondents must be 18 years old or more, and residing in continental France; i.e., excluding the four overseas departments of Guadeloupe, Martinique, French Guiana and Réunion. Those that have smoked more than 100 cigarettes in their life and smoked at least once in the 30 days prior to recruitment were considered to be smokers, whereas the others were considered to be non-smokers.

¹Dept. of Statistics & Actuarial Science, University of Waterloo, Waterloo, Ontario, N2L 3G1

²Data Management Core (DMC) — ITC Project, University of Waterloo

³Dept. of Psychology, University of Waterloo

⁴This document was created using L^AT_EX, and last updated on Aug. 10, 2016

1 SAMPLING DESIGN

1.1 Wave 1

The ITC France Survey follows a random-digit dialling (RDD) sampling design, and its sampling frame is the RDD-B list from Survey Sampling International (SSI). Note that the RDD-B frame includes phone numbers on the white, orange (i.e., phone numbers that are listed, but not commercialized by France Télécom) and red (i.e., unlisted/confidential phone numbers) lists, but households with only cell phones are not included. Furthermore, unlike most ITC surveys in western countries, the ITC France Survey does not use stratification. Hence, households were randomly called nation-wide using RDD until the planned 1700 smokers and 500 non-smokers were interviewed. The quota of 500 non-smokers consisted of 250 non-smokers from non-smoking households, and 250 from households with both smokers and non-smokers. In households with multiple eligible respondents (this includes non-smokers residing with smokers when the corresponding quota was open), the Next Birthday method (Binson et al. (2000)) was used to select a single one. No substitution within household was allowed, except when it was known that the selected respondent would be absent for the entire fieldwork period. When the non-smoker quotas were closed, only households with one or more qualified smokers were deemed eligible.

The ITC France wave 1 sample consists of 1735 adult smokers and of 525 adult non-smokers, for a total of 2260 respondents. This sample of respondents is referred to as cohort 1. Fieldwork was conducted by l’Institute Atoo in Paris.

1.2 Wave 2

Out of the 2260 wave 1 respondents, 1231 smokers and 414 non-smokers were successfully recontacted at wave 2; yielding a retention rate of 72.8% (71.0% for smokers and 78.9% for non-smokers). To compensate for the attrition, 574 additional respondents (473 smokers and 101 non-smokers) were randomly sampled and interviewed; for a total of 2219 respondents (1704 smokers and 515 non-smokers) interviewed at wave 2.

This sample of 574 respondents is referred to as the wave 2 replenishment sample or cohort 2. As in other ITC surveys, replenishment for wave 2 of the ITC France Survey was carried out using the same sampling design and interview protocol as in wave 1 (see section 1.1). Hence, the replenishment sample is representative of the population at the time of data collection (i.e., Sept.–Dec. 2008), rather than those lost to follow-up/attrition. Fieldwork was conducted by l’Institut des Sondages Lavialle (ISL), now GfK-ISL.

1.3 Wave 3

Out of the 2219 wave 2 respondents, 1215 smokers and 373 non-smokers were successfully recontacted at wave 3; yielding a retention rate of 71.6% (71.3% for smokers and 72.4% for non-smokers). Note that of the 373 non-smokers recontacted, 17 had started smoking by wave 3. To compensate for the attrition, 616 additional respondents (502 smokers and 114 non-smokers) were randomly sampled and interviewed; for a total of 2204 respondents (1700 smokers and 504 non-smokers) interviewed at wave 3.

This sample of 616 respondents is referred to as the wave 3 replenishment sample or cohort 3. As in wave 2 and other ITC surveys, replenishment for wave 3 of the ITC France Survey was carried out using the same sampling design and interview protocol as in wave 1 (see section 1.1). Hence, the replenishment sample is representative of the population at the time of data collection (i.e., Sept. – Dec. 2012), rather than those lost to follow-up/attrition. Wave 3 fieldwork was conducted by BVA - Inference Operations.

Figure 1 show the attrition and replenishment of the ITC France Survey samples over all three waves.

2 Weight construction

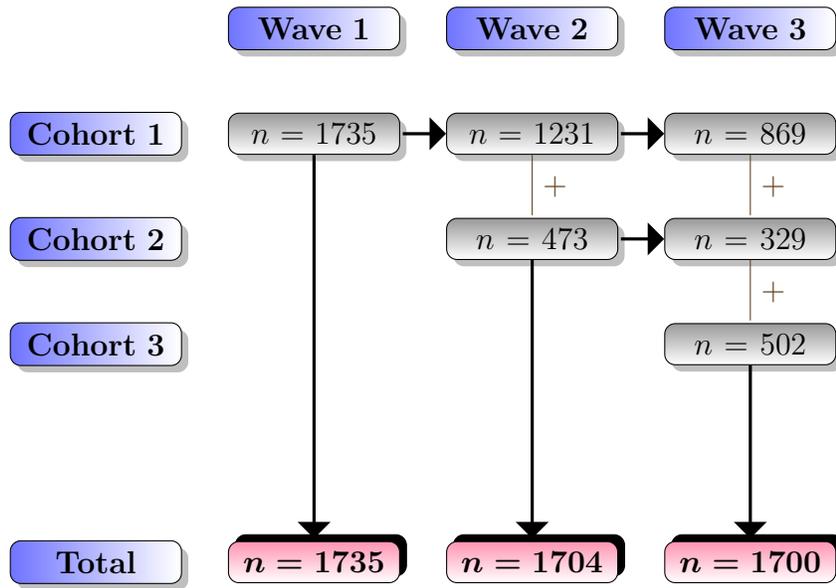
2.1 General comments about weight construction

As with most survey weights, the ITC France sampling weights are constructed to correct and adjust for sample mis-representation caused by unequal sampling probabilities, frame error (i.e., under-coverage and multiplicity) and non-response, as well as improving precision of estimates through the use of auxiliary information (e.g., smoking prevalences). In addition, conservative weight trimming was performed to prevent extreme weight variation arising from a few respondents having very large sampling weights. We briefly describe these key concepts of weight construction in this section, but refer the reader to Levy & Lemeshow (2008), chapter 16, for more detailed information.

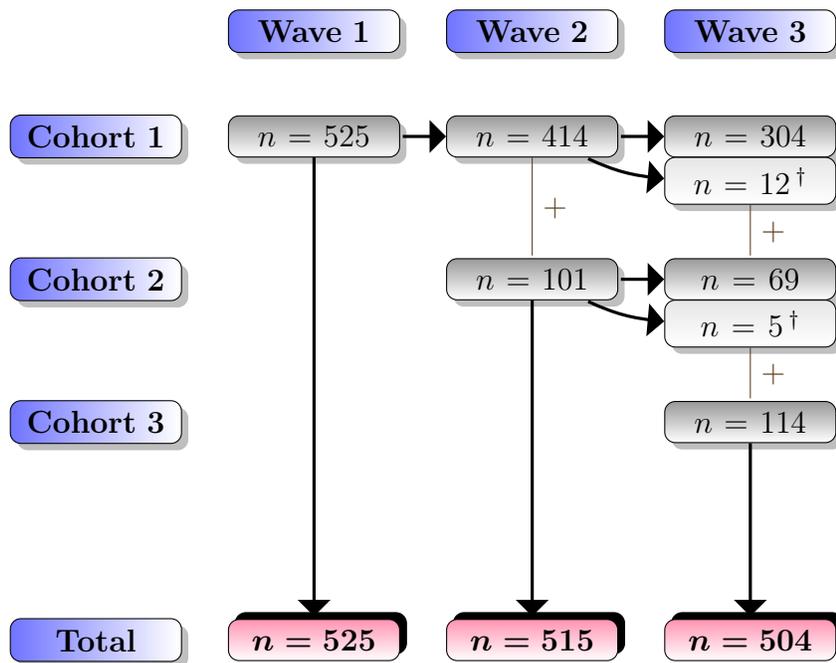
At their base, sampling weights are defined as the inverse of selection probabilities, and thus adjust for sample mis-representation caused by unequal sampling probabilities. For example, a smoker residing alone has a probability of selection twice that of a smoker residing with another smoker.

Frame multiplicity refers to members of the population listed twice or more on the sampling frame. For example, a household with two phone lines has twice the probability of being selected than a household with a single phone line. To correct such bias, respondents to the ITC France Survey were asked how many landlines they have, and their sampling weights were adjusted accordingly; see step 2 of section 2.2.1.

2 WEIGHT CONSTRUCTION



(a) Smokers cohort



[†] Non-smokers who start smoking

(b) Non-Smokers cohort

Figure 1: Attrition and replenishment in the ITC France Survey.

Great efforts are made to create a complete/perfect sampling frame (i.e., a frame that includes all members of the target population, without duplicate and without any erroneous inclusions¹). However, this is seldomly achieved and, consequently, some members of the target population are not included in the sampling frame (i.e., have a 0 probability of being selected). This is known as frame under-coverage, and can result in non-coverage bias. To reduce non-coverage bias in the ITC France Survey, post-stratification adjustments were performed on the sampling weights to ensure that, for each region and age/sex group, the totals of the sampling weights equal known benchmarks; see steps 3 and 4 in section 2.2.1 for the smoker weights, and steps 3 and 4 in section 2.2.2 for the non-smoker weights. Note that these benchmark figures are also referred to as calibration or target figures, and thus this post-stratification adjustment is also referred to as weight calibration.

If non-respondents behave differently than respondents, then inference based solely on the sample of respondents will be biased unless adjustments are made. The greater the proportion of non-response, the greater this bias can be. In the ITC France Survey, the post-stratification adjustment described in the above paragraph also adjust for non-coverage bias. It should be noted that if data are missing completely at random (MCAR, see Little & Rubin (2002)) within each region and age/sex group, then non-response bias will be completely eliminated. Realistically though, non-response bias is greatly reduced, but not eliminated in the ITC France Survey.

The distribution of sampling weights is often skewed to the right; echoing the fact that most populations are composed of many average/typical members and of few atypical ones. Average members have a fairly high probability of selection, and thus most sampling weights are quite small. There are however few members of the population that have a much smaller probability of selection, and consequently have sampling weights that are quite large. These few large weights can be the source of high weight variation, which increases the variability of estimators and thus decreases precision. To correct for this, large weights are often trimmed in the weight construction process. This must be done with care and conservatively, as trimming can increase bias. There are various ways of trimming sampling weights. In the ITC France Survey, trimming was done by capping the number of adults (and thus the number of smokers) in each household at 4 (see step 1 in section 2.2.1). Capping is a fairly conservative weight trimming technique and, since it is done at the beginning of weight construction, helps minimize potentially biasing estimates.

It is well known from survey sampling theory that, in the vast majority of cases, the ratio estimator has much greater precision than the commonly used Horvitz-Thompson estimator. Heuristically, this is due to the fact that the ratio estimator utilizes auxiliary

¹Erroneous inclusions refers to units that are not part of the target population, but included in the sampling frame.

2 WEIGHT CONSTRUCTION

(i.e., additional) information in addition to the sampling weights; whereas the Horvitz-Thompson estimator does not. As mentioned above however, smoking prevalence figures were used to calibrate the ITC France sampling weights in order to reduce biases from frame errors and non-response. Our calibrating procedure yields (so-called) ratio weights, which enable all estimators to inherit the increased precision of the ratio estimator.

All sampling weights for the ITC France Survey were computed using the statistical software R (<http://www.r-project.org>).

2.2 Wave 1 weights

Two sets of weights were computed at wave 1:

- i) Section 2.2.1 describes the computation of the **cross-sectional wave 1 weights for smokers** for the 1735 smokers who completed the wave 1 survey.
- ii) Section 2.2.2 describes the computation of the **cross-sectional wave 1 weights for non-smokers** for the 525 non-smokers who completed the wave 1 survey.

Since no respondent can have both a smoker and a non-smoker wave 1 weight, both sets were combined into a single variable, labelled `aDE41919v`.

It should be noted that the smoker weights were calibrated to smoking prevalence (see step 4 of section 2.2.1) and rescaled to have a mean equal to 1 (see step 5 of section 2.2.1). Similarly, the non-smoker weights were calibrated to non-smoking prevalence and rescaled to have a mean equal to 1. Consequently, these weights should not be used to estimate population totals (e.g., the total number of daily smokers). However, all of them can obviously be used to estimate population means and proportions/percentages, as well as in various statistical models (e.g., logistic and linear regressions). Un-rescaled weights (i.e., prior to step 5 of section 2.2.1) are also available for both smokers and non-smokers; see variable `aDE41915v`. Because these un-rescaled weights sum to the number of adult smokers (or non-smokers) in France, they should not be used when doing multi-country analyses.

2.2.1 Smoker weights

Computation of sampling weights for the 1735 smokers who completed the wave 1 survey proceeded as follows:

Step 1: Each respondent was first assigned an initial weight $w_i^{(1)}$, which can be viewed as an adjustment for the probability of selection within a given household while

the non-smoker quotas were open and after they were closed. Formally, these $w_i^{(1)}$ weights are given by

$$w_i^{(1)} = \frac{\#\text{smokers}_i \times \#\text{adults}_i}{\#\text{smokers}_i \times 0.1048 + \#\text{adults}_i \times 0.1621}$$

where i stands for the i^{th} respondent, $\#\text{smokers}_i$ is the number of adult smokers in the household, and $\#\text{adults}_i$ is the number of adults in the household. The above formula for $w_i^{(1)}$ accounts for the probability of selecting a smoker in a household with smokers only and in a household with both smokers and non-smokers, as well as the probability of selecting such households while the non-smoker quotas were open and after they were closed; see appendix for details. Recall that $\#\text{adults}_i$ was capped at 4 to prevent large households from having undue influence on the weights; thus, $\#\text{smokers}_i \leq \#\text{adults}_i \leq 4$.

Step 2: The $w_i^{(1)}$ weights were then multiplied by 1 if the corresponding household had a single personal phone line and by 1/2 if it had two or more such lines. The resulting weights are labelled $w_i^{(2)}$, and are formally defined as

$$w_i^{(2)} = w_i^{(1)} \times \frac{1}{1 + I(\#\text{lines}_i > 1)}$$

where $I(\#\text{lines}_i > 1) = 1$ if the household of the i^{th} respondent has two or more personal phone lines, and 0 otherwise.

Step 3: A post-stratification adjustment was then performed to calibrate the $w_i^{(2)}$ weights to known proportions of France adult population residing in each of the 8 Union des Annonceurs (UDA) regions, as of Nov. 2006; see table A.1, column 2, and figure 2. For respondents residing in UDA region k , this consisted in multiplying their $w_i^{(2)}$ weights by a factor f_k to produce adjusted $w_i^{(3)}$ weights. These $w_i^{(3)}$ weights are such that their sum over all respondents residing in UDA region k divided by their sum over all 1735 respondents is equal to the known proportion in question. In other words,

$$w_i^{(3)} = w_i^{(2)} \times f_k$$

where

$$f_k = \frac{p_k}{\sum_{k \in F_k} w_i^{(2)} / \sum_{i=1}^n w_i^{(2)}}$$

and where p_1, \dots, p_8 are given in column 2 of table A.1 and F_k is the set of all respondents residing in UDA region k . This was done to compensate for differentials achieved sampling fractions from region to region.

2 WEIGHT CONSTRUCTION



Figure 2: Union des Annonceurs (UDA) regions.
(Bassin parisien est & ouest were combined for weight calculations)

Step 4: The $w_i^{(3)}$ weights were then calibrated to smoking prevalence by age/sex groups using the same post-stratification technique used in step 3. To this end, age was divided into 5 intervals (i.e., [18,25), [25,35), [35,50), [50,60) and [60,100)). The resulting $w_i^{(4)}$ weights thus sum to the estimated number of adults smokers in each of the 10 age/sex cells of table A.2. Formally,

$$w_i^{(4)} = w_i^{(3)} \times \frac{c_\ell}{\sum_{i \in C_\ell} w_i^{(3)}}$$

where c_1, \dots, c_{10} are given in column 3 of table A.2 and C_ℓ is the set of all respondents in cell ℓ . The calibration figures of table A.2 were obtained by combining population estimates (as of Jan. 1, 2005) from INSEE to prevalence estimates from Beck et al. (2005); see appendix for details.

Step 5: To facilitate comparisons across multiple ITC countries, the $w_i^{(4)}$ weights were rescaled to have a mean equal to 1 or, equivalently, to sum to $n = 1735$ (the number of smokers who completed the wave 1 survey). This yielded the $w_i^{(5)}$

weights, which are formally defined as

$$w_i^{(5)} = w_i^{(4)} \times \text{RF}$$

where RF is a rescaling factor and is given by

$$\text{RF} = \frac{n}{\sum_{i=1}^n w_i^{(4)}} = \frac{1735}{\sum_{i=1}^n w_i^{(4)}} .$$

Note: the coefficient of variation (cv) of the $w_i^{(5)}$ weights is 0.455.

2.2.2 Non-smoker weights

Computation of sampling weights for the 525 non-smokers who completed the wave 1 survey proceeded alike for the smokers' weights; i.e.,

Step 1: Each respondent was first assigned an initial weight $w_i^{(1)}$, which can be viewed as an adjustment for the probability of selection within a given household while the non-smoker quotas were open. Using the same notation as in step 1 of section 2.2.1, these $w_i^{(1)}$ weights are formally given by

$$w_i^{(1)} = \#\text{adults}_i \times 7.38$$

The multiplicative term 7.38 adjusts for the probability of selecting a non-smoker in a household with non-smokers only and in a household with both smokers and non-smokers, as well as the probability of selecting such households while both non-smoker quotas were open and after the non-smokers in non-smoking households quota was closed; see appendix for details.

Step 2: The $w_i^{(1)}$ weights were multiplied by 1 or 1/2 to adjust for the number of personal phone lines in the household; same as step 2 of section 2.2.1.

Step 3: A post-stratification calibration was performed to compensate for differential achieved sampling fractions from UDA region to UDA region, same as step 3 of section 2.2.1.

Step 4: The weights were then calibrated to non-smoking prevalence by age/sex groups. This was done the same way as in step 4 of section 2.2.1 with the exception that the calibration figures for non-smokers were used instead of those for smokers. Hence, the resulting weights sum to the estimated number of adult non-smokers in each of the 10 age/sex cells of table A.2. These non-smoking calibration figures were obtained by combining the same two surveys as in step 4 of section 2.2.1, and by simply taking one minus the smoking prevalence figure as the corresponding non-smoking prevalence figure; see appendix for details.

2 WEIGHT CONSTRUCTION

Step 5: The weights computed in step 4 above were rescaled to sum to sample size $n = 525$, same as step 5 of section 2.2.1.

Note: the coefficient of variation (cv) of the wave 1 non-smoker weights is 0.541.

2.3 Wave 2 weights

Six sets of weights were computed at wave 2.

Section 2.3.1 describes the computations of the smoker weights:

- i) **Waves 1–2 longitudinal weights for smokers** were computed for the 1231 smokers recruited at wave 1 that were retained and interviewed at wave 2. Note that 164 of these respondents had quit smoking at wave 2. Explicitly, in the ITC France Survey and the majorities of ITC surveys, if a respondent is recruited as a smoker, he/she will always be considered as a smoker when computing his/her weights, regardless of his/her current smoking status.
- ii) **Wave 2 cross-sectional weights for cohort smokers** were computed for the above 1231 respondents.
- iii) **Wave 2 cross-sectional weights for newly recruited smokers** were computed for the 473 smokers recruited and interviewed at wave 2.

Section 2.3.2 describes the computations of the non-smoker weights:

- iv) **Waves 1–2 longitudinal weights for non-smokers** were computed for the 414 non-smokers recruited at wave 1 that were retained and interviewed at wave 2.
- v) **Wave 2 cross-sectional weights for cohort non-smokers** were computed for the above 414 respondents.
- vi) **Wave 2 cross-sectional weights for newly recruited non-smokers** were computed for the 101 non-smokers recruited and interviewed at wave 2.

The two sets of longitudinal weights were constructed to adjust for attrition between waves 1 and 2, thus ensuring that the subset of respondents who completed both waves still represents the population at the time of wave 1 (i.e., Dec. 2006–Feb. 2007). Hence, these wave 2 longitudinal weights were calibrated using wave 1 figures (i.e., table A.1, column 2, and table A.2). The two sets of cross-sectional weights for cohort respondents were constructed for the same subset of respondents who completed both waves. However, new calibration figures (i.e., table A.1, column 3, and table A.3) were used to ensure that these respondents represent the population at the time of wave 2 (i.e., Sept.–Dec. 2008).

Lastly, cross-sectional weights were constructed for the new smokers and non-smokers recruited at wave 2.

Since no respondent can have both a smoker and a non-smoker longitudinal weights, both sets of longitudinal weights were combined into a single variable, labelled `bDE41921v`. Similarly, all four sets of cross-sectional weights were combined into a single variable, labelled `bDE41919v`. All ITC France wave 2 weights were calibrated to smoking prevalence by age/sex group and rescaled to have a mean equal to 1. Consequently, these weights should not be used to estimate population totals (e.g., the total number of daily smokers).

2.3.1 Smoker weights

Starting with $w_i^{(0)}$, the wave 1 smoker weight for the i^{th} respondent (computed in section 2.2.1), computation of the 1231 **waves 1–2 longitudinal weights for smokers** proceeded as follows:

Step 1: The $w_i^{(0)}$ weights were first re-calibrated to the proportions of adults residing in each of the 8 UDA regions as of Nov. 2006; see column 2 of table A.1. This was done the same way as in step 3 of section 2.2.1.

Step 2: The weights were then re-calibrated to smoking prevalence by age/sex groups. This was done the same way as step 4 of section 2.2.1 and using the same prevalence figures (i.e., column 3 of table A.2).

Step 3: Lastly, the weights were rescaled to sum to sample size $n = 1231$, same as step 5 of section 2.2.1.

Note: the coefficient of variation (cv) of the waves 1–2 longitudinal weights for smokers is 0.488.

Starting with $w_i^{(0)}$, the wave 1 smoker weight for the i^{th} respondent (computed in section 2.2.1), computation of the 1231 **wave 2 cross-sectional weights for cohort smokers** proceeded as follows:

Step 1: If the i^{th} respondent is one of the 34 smokers who moved to a new UDA region between waves 1 and 2, his/her $w_i^{(0)}$ was replaced by the average of the wave 1 smoker weights of respondents in the same age/sex group living in his/her new UDA region.

Step 2: The $w_i^{(0)}$ weights were calibrated to the proportions of adults residing in each of the 8 UDA regions, but using the updated figures given in column 3 of table A.1 instead of those in column 2. This was done the same way as in step 3 of section 2.2.1.

2 WEIGHT CONSTRUCTION

Step 3: The weights were then calibrated to smoking prevalence by age/sex groups. This was done the same way as in step 4 of section 2.2.1, but using the updated figures given in table A.3 instead of those in table A.2.

Step 4: Lastly, the weights were rescaled to sum to sample size $n = 1231$, same as step 5 of section 2.2.1.

Note: the coefficient of variation (cv) of the wave 2 cross-sectional weights for cohort smokers is 0.488.

Computation of the 473 **wave 2 cross-sectional weights for newly recruited smokers** proceeded as follows:

Step 1: As in section 2.2.1, each new respondent was first assigned an initial weight $w_i^{(1)}$, which can be viewed as an adjustment for the probability of selection within a given household while the non-smoker quotas were open and after they were closed. Formally, these $w_i^{(1)}$ weights are given by

$$w_i^{(1)} = \frac{\#\text{smokers}_i \times \#\text{adults}_i}{\#\text{smokers}_i \times 0.0282 + \#\text{adults}_i \times 0.2485}$$

where i stands for the i^{th} respondent, $\#\text{smokers}_i$ is the number of adult smokers in the household, and $\#\text{adults}_i$ is the number of adults in the household. The above formula for $w_i^{(1)}$ accounts for the probability of selecting a smoker in a household with smokers only and in a household with both smokers and non-smokers, as well as the probability of selecting such households while the non-smoker quotas were open and after they were closed; see appendix for details. Recall that $\#\text{adults}_i$ was capped at 4 to prevent large households from having undue influence on the weights; thus, $\#\text{smokers}_i \leq \#\text{adults}_i \leq 4$.

Step 2: The $w_i^{(1)}$ weights were multiplied by 1 or 1/2 to adjust for the number of personal phone lines in the household; same as step 2 of section 2.2.1.

Step 3: The weights were calibrated to the proportions of adults residing in each of the 8 UDA regions. This was done the same way as step 3 of section 2.2.1, but using the wave 2 figures in column 3 of table A.1 instead of wave 1 figures in column 2.

Step 4: The weights were then calibrated to smoking prevalence by age/sex groups. This was done the same way as step 4 of section 2.2.1, but using the wave 2 prevalence figures of table A.3 instead of wave 1 figures.

Step 5: Lastly, the weights were rescaled to sum to sample size $n = 473$, same as step 5 of section 2.2.1.

Note: the coefficient of variation (cv) of the wave 2 cross-sectional weights for newly recruited smokers is 0.661.

2.3.2 Non-smoker weights

Starting with $w_i^{(0)}$, the wave 1 non-smoker weight for the i^{th} respondent (computed in section 2.2.2), computation of the 414 **waves 1–2 longitudinal weights for non-smokers** proceeded as follows:

Step 1: The $w_i^{(0)}$ weights were first re-calibrated to the proportions of adults residing in each of the 8 UDA regions as of Nov. 2006; see column 2 of table A.1. This was done the same way as in step 3 of section 2.2.1.

Step 2: The weights were then re-calibrated to non-smoking prevalence by age/sex groups. This was done the same way as in step 4 of section 2.2.2 and using the same prevalence figures (i.e., column 4 of table A.2).

Step 3: Lastly, the weights were rescaled to sum to sample size $n = 414$, same as step 5 of section 2.2.1.

Note: the coefficient of variation (cv) of the waves 1–2 longitudinal weights for non-smokers is 0.587.

Starting with $w_i^{(0)}$, the wave 1 non-smoker weight for the i^{th} respondent (computed in section 2.2.2), computation of the 414 **wave 2 cross-sectional weights for cohort non-smokers** proceeded as follows:

Step 1: If the i^{th} respondent is one of the 3 non-smokers who moved to a new UDA region between waves 1 and 2, his/her $w_i^{(0)}$ was replaced by the average of the wave 1 non-smoker weights of respondents in the same age/sex group living in his/her new UDA region.

Step 2: The $w_i^{(0)}$ weights were calibrated to the proportions of adults residing in each of the 8 UDA regions, but using the updated figures given in column 3 of table A.1 instead of those given in column 2. This was done the same way as in step 3 of section 2.2.1.

Step 3: The weights were then calibrated to non-smoking prevalence by age/sex groups. This was done the same way as in step 4 of section 2.2.2, but using the updated figures given in table A.3 instead of those in table A.2.

Step 4: Lastly, the weights were rescaled to sum to sample size $n = 414$, same as step 5 of section 2.2.1.

2 WEIGHT CONSTRUCTION

Note: the coefficient of variation (cv) of the wave 2 cross-sectional weights for cohort non-smokers is 0.590.

Computation of the 101 **wave 2 cross-sectional weights for newly recruited non-smokers** proceeded as follows:

Step 1: As in section 2.2.2, each new respondent was first assigned an initial weight $w_i^{(1)}$, which can be viewed as an adjustment for the probability of selection within a given household while the non-smoker quotas were open. These $w_i^{(1)}$ weights are formally given by

$$w_i^{(1)} = \#adults_i \times 16.28$$

where i stands for the i^{th} respondent, and $\#adults_i$ is the number of adults residing in that i^{th} household. The multiplicative term 16.28 adjusts for the probability of selecting a non-smoker in a household with non-smokers only and in a household with both smokers and non-smokers, as well as the probability of selecting such households while both non-smoker quotas were open and after the the non-smokers in non-smoking household quota was closed; see appendix for details.

Step 2: The $w_i^{(1)}$ weights were multiplied by 1 or 1/2 to adjust for the number of personal phone lines in the household; same as step 2 of section 2.2.1.

Step 3: The weights were calibrated to the proportions of adults residing in each of the 8 UDA regions. This was done the same way as in step 3 of section 2.2.1, but using the wave 2 figures in column 3 of table A.1 instead of wave 1 figures in column 2.

Step 4: The weights were then calibrated to non-smoking prevalence by age/sex groups. This was done the same way as in step 4 of section 2.2.2, but using the wave 2 prevalence figures of table A.3 instead of wave 1 figures.

Step 5: Lastly, the weights were rescaled to sum to sample size $n = 101$, same as step 5 of section 2.2.1.

Note: the coefficient of variation (cv) of the wave 2 cross-sectional weights for newly recruited non-smokers is 0.642.

2.4 Wave 3 weights

Eight sets of weights were computed at wave 3.

Section 2.4.1 describes the computations of the smoker weights:

- i) **Waves 1–3 longitudinal weights for smokers** were computed for the 869 respondents recruited as smokers at wave 1 that were retained and interviewed at waves 2 and 3. Note that 235 of these respondents had quit smoking by wave 3 (88 quit at wave 2 and 147 quit at wave 3). These 869 smokers also include 40 respondents who quit smoking at wave 2, but started smoking again at wave 3. Explicitly, in the ITC France Survey and the majorities of ITC surveys, if a respondent is recruited as a smoker, he/she will always be considered as a smoker when computing his/her weights, regardless of his/her current smoking status.
- ii) **Waves 2–3 longitudinal weights for smokers** were computed for the 1198 respondents recruited as smokers at either wave 1 or 2 that completed both the wave 2 and 3 surveys. Note that 297 of these respondents had quit smoking by wave 3 (88 quit at wave 2 and 209 quit at wave 3); see above paragraph.
- iii) **Wave 3 cross-sectional smoker weights for cohort respondents** were computed for the 1215 respondents that smoked (or had quit smoking) when interviewed at wave 3. These 1215 respondents consist of the 1198 cohort smokers mentioned above, and of 17 respondents that were recruited as non-smokers at waves 1 or 2, but started smoking by wave 3 (12 were recruited at wave 1, and 5 were recruited at wave 2). Hence, if a respondent was recruited as a non-smoker, but had started smoking by wave 3, he/she will receive a cross-sectional smoker weight at wave 3.
- iv) **Wave 3 cross-sectional weights for newly recruited smokers** were computed for the 502 smokers recruited and interviewed at wave 3.

Section 2.4.2 describes the computations of the non-smoker weights:

- v) **Waves 1–3 longitudinal weights for non-smokers** were computed for the 304 respondents recruited as non-smokers at wave 1 that were retained and interviewed at waves 2 and 3, as well as remaining non-smoker. Hence, the 12 non-smokers recruited at wave 1, interviewed at waves 2 and 3, but who had started smoking by wave 3 are not part of this group. Unlike smokers who quit (see item i above), no longitudinal weights will be computed for non-smokers who start smoking. However, cross-sectional smoker weights (see item iii above) will be computed for such individuals.
- vi) **Waves 2–3 longitudinal weights for non-smokers** were computed for the 373 respondents recruited as non-smokers at waves 1 or 2 that completed both the wave 2 and 3 surveys, as well as remaining non-smoker. Hence, the 17 non-smokers who had started smoking by wave 3 are not part of this group. As mentioned above, no longitudinal weights will be computed for non-smokers who start smoking. However, cross-sectional smoker weights (see item iii above) will be computed for such individuals.

2 WEIGHT CONSTRUCTION

- vii) **Wave 3 cross-sectional non-smoker weights for cohort respondents** were computed for the 373 respondents recruited as non-smokers at waves 1 or 2 that completed both the wave 2 and 3 surveys and had not started smoking at the time of wave 3 fieldwork. Hence, no wave 3 cross-sectional non-smoker weights for cohort respondents were computed for the 17 respondents that had started smoking by wave 3. Instead, wave 3 cross-sectional smoker weights for cohort respondents were computed for these 17 respondents; see item iii above.
- viii) **Wave 3 cross-sectional weights for newly recruited non-smokers** were computed for the 114 non-smokers recruited and interviewed at wave 3.

The two sets of waves 1–3 longitudinal weights were constructed to adjust for attrition between waves 1 and 3, thus ensuring that the subset of respondents who completed all three waves still represents the population at the time of wave 1 (i.e., Dec. 2006–Feb. 2007). Similarly, the two sets of waves 2–3 longitudinal weights were constructed to adjust for attrition between waves 2 and 3, thus ensuring that the subset of respondents who completed both waves still represent the population at the time of wave 2 (i.e., Sept.–Dec. 2008). Hence, the waves 1–3 longitudinal weights were calibrated using wave 1 figures (i.e., table A.1, column 2, and table A.2), and the waves 2–3 longitudinal weights were calibrated using wave 2 figures (i.e., table A.1, column 3, and table A.3). The two sets of cross-sectional weights for cohort respondents were constructed for the same subset of respondents who completed wave 2. However, new calibration figures (i.e., table A.1, column 4, and table A.4) were used to ensure that these respondents represent the population at the time of wave 3 (i.e., Sept.–Dec. 2012). Lastly, cross-sectional weights were constructed for the new smokers and non-smokers recruited at wave 3.

Since no respondent can have both a smoker and a non-smoker longitudinal weight, both sets of waves 1–3 longitudinal weights and waves 2–3 longitudinal weights were combined into a single variable for each, labelled `cDE41921v` and `cDE41923v`, respectively. Similarly, all four sets of cross-sectional weights were combined into a single variable, labelled `cDE41919v`. All ITC France wave 3 weights were calibrated to smoking prevalence by age/sex group and rescaled to have a mean equal to 1. Consequently, these weights should not be used to estimate population totals (e.g., the total number of daily smokers).

2.4.1 Smoker weights

Starting with $w_i^{(0)}$, the wave 1–2 longitudinal smoker weight for the i^{th} respondent (computed in section 2.3.1), computation of the 869 **waves 1–3 longitudinal weights for smokers** proceeded as follows:

Step 1: The $w_i^{(0)}$ weights were first re-calibrated to the proportions of adults residing in each of the 8 UDA regions as of Nov. 2006; see column 2 of table A.1. This was done the same way as in step 3 of section 2.2.1.

Step 2: The weights were then re-calibrated to smoking prevalence by age/sex groups. This was done the same way as step 4 of section 2.2.1 and using the same prevalence figures (i.e., column 3 of table A.2).

Step 3: Lastly, the weights were rescaled to sum to sample size $n = 869$, same as step 5 of section 2.2.1.

Note: the coefficient of variation (cv) of the waves 1–3 longitudinal weights for smokers is 0.512.

The calculation of the 1198 **waves 2–3 longitudinal weights for smokers** were computed in the same way as described above for the waves 1–3 longitudinal weights for smokers with the exception that the wave 2 cross-sectional smoker weight for the i^{th} respondent was used as the starting weight, $w_i^{(0)}$ (computed in section 2.3.1). The weights were also re-calibrated to the proportion of adults residing in each of the 8 UDA regions, as of Aug. 2008 in order to represent the population at the time of wave 2; see column 3 of table A.1. The re-calibration to smoking prevalence by age/sex groups also used the updated figures provided in table A.3. Note: the coefficient of variation (cv) of the waves 2–3 longitudinal weights for smokers is 0.575.

Starting with $w_i^{(0)}$, the wave 2 smoker weight for the i^{th} respondent (computed in section 2.3.1), computation of the 1215 **wave 3 cross-sectional smoker weights for cohort respondents** proceeded as follows:

Step 1: If the i^{th} respondent is one of the 17 non-smokers who became smokers between waves 2 and 3, his/her $w_i^{(0)}$ was replaced by the average of the wave 2 smoker weights of respondents in the same age/sex group living in his/her UDA region.

Step 2: The $w_i^{(0)}$ weights were calibrated to the proportions of adults residing in each of the 8 UDA regions using the figures given in column 4 of table A.1. This was done the same way as in step 3 of section 2.2.1.

Step 3: The weights were then calibrated to smoking prevalence by age/sex groups. This was done the same way as in step 4 of section 2.2.1, but using the updated figures given in table A.4 instead of those in table A.2.

Step 4: Lastly, the weights were rescaled to sum to sample size $n = 1215$, same as step 5 of section 2.2.1.

Note: the coefficient of variation (cv) of the wave 3 cross-sectional weights for cohort smokers is 0.542.

2 WEIGHT CONSTRUCTION

Computation of the 502 **wave 3 cross-sectional weights for newly recruited smokers** proceeded as follows:

Step 1: As in section 2.2.1, each new respondent was first assigned an initial weight $w_i^{(1)}$, which can be viewed as an adjustment for the probability of selection within a given household while the non-smoker quotas were opened and after they were closed. Formally, these $w_i^{(1)}$ weights are given by

$$w_i^{(1)} = \frac{\#\text{smokers}_i \times \#\text{adults}_i}{\#\text{smokers}_i \times 0.0146 + \#\text{adults}_i \times 0.3463}$$

where i stands for the i^{th} respondent, $\#\text{smokers}_i$ is the number of adult smokers in the household, and $\#\text{adults}_i$ is the number of adults in the household. The above formula for $w_i^{(1)}$ accounts for the probability of selecting a smoker in a household with smokers only and in a household with both smokers and non-smokers, as well as the probability of selecting such households while the non-smoker quotas were open and after they were closed; see appendix for details. Recall that $\#\text{adults}_i$ was capped at 4 to prevent large households from having undue influence on the weights; thus, $\#\text{smokers}_i \leq \#\text{adults}_i \leq 4$.

Step 2: The weights were calibrated to the proportions of adults residing in each of the 8 UDA regions. This was done the same way as step 3 of section 2.2.1, but using the wave 3 figures in column 4 of table A.1 instead of the wave 1 figures in column 2.

Step 3: The weights were then calibrated to smoking prevalence by age/sex groups. This was done the same way as step 4 of section 2.2.1, but using the wave 3 prevalence figures of table A.4 instead of the wave 1 figures of table A.2.

Step 4: Lastly, the weights were rescaled to sum to sample size $n = 502$, same as step 5 of section 2.2.1.

Note: the coefficient of variation (cv) of the wave 3 cross-sectional weights for newly recruited smokers is 0.636.

2.4.2 Non-smoker weights

Starting with $w_i^{(0)}$, the wave 1–2 longitudinal non-smoker weight for the i^{th} respondent (computed in section 2.3.2), computation of the 304 **waves 1–3 longitudinal weights for non-smokers** proceeded as follows:

Step 1: The $w_i^{(0)}$ weights were first re-calibrated to the proportions of adults residing in each of the 8 UDA regions as of Nov. 2006; see column 2 of table A.1. This was done the same way as in step 3 of section 2.2.1.

Step 2: The weights were then re-calibrated to non-smoking prevalence by age/sex groups. This was done the same way as in step 4 of section 2.2.2 and using the same prevalence figures (i.e., column 4 of table A.2).

Step 3: Lastly, the weights were rescaled to sum to sample size $n = 304$, same as step 5 of section 2.2.1.

Note: the coefficient of variation (cv) of the waves 1–3 longitudinal weights for non-smokers is 0.659.

The calculation of the 373 **waves 2–3 longitudinal weights for non-smokers** were computed in the same way as described above for the waves 1–3 longitudinal weights for non-smokers with the exception that the wave 2 cross-sectional non-smoker weight for the i^{th} respondent was used as the starting weight, $w_i^{(0)}$ (computed in section 2.3.2). The weights were also re-calibrated to the proportion of adults residing in each of the 8 UDA regions, as of Aug. 2008 in order to represent the population at the time of wave 2; see column 3 of table A.1. The re-calibration to non-smoking prevalence by age/sex groups also used the updated figures provided in table A.3. Note: the coefficient of variation (cv) of the waves 2–3 longitudinal weights for smokers is 0.686.

Starting with $w_i^{(0)}$, the wave 2 non-smoker weight for the i^{th} respondent (computed in section 2.3.2), computation of the 373 **wave 3 cross-sectional non-smoker weights for cohort respondents** proceeded as follows:

Step 1: The $w_i^{(0)}$ weights were calibrated to the proportions of adults residing in each of the 8 UDA regions using the figures given in column 4 of table A.1. This was done the same way as in step 3 of section 2.2.1.

Step 2: The weights were then calibrated to non-smoking prevalence by age/sex groups. This was done the same way as in step 4 of section 2.2.2, but using the figures given in table A.4 instead of those in table A.2.

Step 3: Lastly, the weights were rescaled to sum to sample size $n = 373$, same as step 5 of section 2.2.1.

Note: the coefficient of variation (cv) of the wave 3 cross-sectional weights for cohort non-smokers is 0.733.

Computation of the 114 **wave 3 cross-sectional weights for newly recruited non-smokers** proceeded as follows:

2 WEIGHT CONSTRUCTION

Step 1: As in section 2.2.2, each new respondent was first assigned an initial weight $w_i^{(1)}$, which can be viewed as an adjustment for the probability of selection within a given household while the non-smoker quotas were open. These $w_i^{(1)}$ weights are formally given by

$$w_i^{(1)} = \#adults_i \times 12.48$$

where i stands for the i^{th} respondent, and $\#adults_i$ is the number of adults residing in that i^{th} household. The multiplicative term 12.48 adjusts for the probability of selecting a non-smoker in a household with non-smokers only and in a household with both smokers and non-smokers, as well as the probability of selecting such households while both non-smoker quotas were open and after the the non-smokers in non-smoking household quota was closed; see appendix for details.

Step 2: The weights were calibrated to the proportions of adults residing in each of the 8 UDA regions. This was done the same way as in step 3 of section 2.2.1, with the exception that two UDA regions, Nord and Est, were collapsed; see note at foot of table A.4. The wave 3 figures in column 4 of table A.1 were also used instead of the wave 1 figures in column 2.

Step 3: The weights were then calibrated to non-smoking prevalence by age/sex groups. This was done the same way as in step 4 of section 2.2.2, with the exception that for females, age was collapsed into four intervals, and for males, age was collapsed into two intervals; see note at foot of table A.4. The wave 3 prevalence figures of table A.4 were also used instead of the wave 1 figures.

Step 4: Lastly, the weights were rescaled to sum to sample size $n = 114$, same as step 5 of section 2.2.1.

Note: the coefficient of variation (cv) of the wave 3 cross-sectional weights for newly recruited non-smokers is 0.752.

Acknowledgements

Core funding for the ITC Project is provided by the U.S. National Cancer Institute (NCI) to the Roswell Park TTURC (P50 CA111236 & P01 CA138389), the Canadian Institutes of Health Research (CIHR; Grant #79551 & #115016), and by the Ontario Institute for Cancer Research. Major funding for the ITC France Survey provided by l'Institut national de prévention et d'éducation pour la santé (INPES), l'Observatoire français des drogues et des toxicomanies (OFDT) and l'Institut national du cancer (INCa).

References

- Beck, F., Gautier, A., Guinard, R. & Richard, J.-B., eds (2013), *Baromètre santé 2010: Attitudes et comportements de santé*, Institut national de prévention et d'éducation pour la santé (INPES), Saint-Denis, France.
- Beck, F., Guilbert, P. & Gautier, A., eds (2005), *Baromètre santé 2005: Attitudes et comportements de santé*, Institut national de prévention et d'éducation pour la santé (INPES), Saint-Denis, France.
- Binson, D., Canchola, J. A. & Catania, J. A. (2000), 'Random selection in a national telephone survey: a comparison of the kish, next-birthday and last-birthday methods', *Journal of Official Statistics* **16**, 53–60.
- INSEE (2005), 'Estimations de population au 1^{er} janvier 2005', Saint-Denis, France.
- INSEE (2008), 'Estimations de population au 1^{er} janvier 2008', Saint-Denis, France.
- INSEE (2010), 'Estimations de population au 1^{er} janvier 2010', Saint-Denis, France.
- Levy, P. S. & Lemeshow, S. (2008), *Sampling of Populations: Methods and Applications*, 4 edn, John Wiley & Sons, Hoboken, NJ.
- Little, R. J. A. & Rubin, D. B. (2002), *Statistical Analysis with Missing Data*, 2 edn, John Wiley & Sons, Hoboken, NJ.

A Benchmark/calibration figures

The estimated number of smokers and non-smokers given in tables A.2 and A.3 were obtained by combining population estimates from l’Institut national de la statistique et des études économiques (INSEE) to prevalence estimates from Beck et al. (2005). The estimated number of smokers and non-smokers given in table A.4 were obtained by combining population estimates from l’Institut national de la statistique et des études économiques (INSEE) to prevalence estimates from Beck et al. (2013). To estimate the number of smokers, population estimates for a given age/sex/UDA region combination were simply multiplied by the smoking prevalence for the same age/sex/UDA region combination. The figures were then collapsed over the 8 UDA regions. Estimation of the number of non-smokers proceeded the same way, except that population estimates were multiplied by one minus the smoking prevalence of the corresponding age/sex/UDA region combination. Population estimates as of Jan. 1, 2005 were used for table A.2, whereas figures as of Jan. 1, 2008 were used for table A.3, and figures as of Jan. 1, 2010 were used for table A.4.

Régions UDA	Proportion (%)		
	as of Nov. 2006	as of Aug. 2008	as of Aug. 2012
Bassin parisien	17.7	17.6	18.1
Région parisienne	18.5	17.4	16.8
Centre est	11.9	11.9	12.1
Est	8.8	8.7	9.2 ^a
Méditerranée	12.3	12.8	11.6
Nord	6.6	6.6	6.7 ^a
Ouest	13.3	14.1	14.3
Sud ouest	10.9	10.9	11.2
Total	100	100	100

Cells sharing the same letter were collapsed when computing the wave 3 cross-sectional weights for newly recruited smokers.

Table A.1: France adult (i.e., 18+) population by UDA region.

Sex	Age	#smokers	#non-smokers
male	[18,25)	1,282,253	1,517,111
male	[25,35)	1,891,016	2,159,529
male	[35,50)	2,550,602	3,829,689
male	[50,60)	1,152,358	2,822,679
male	[60,100)	760,680	4,494,062
female	[18,25)	1,115,512	1,614,766
female	[25,35)	1,466,115	2,554,555
female	[35,50)	2,216,570	4,328,593
female	[50,60)	790,548	3,292,319
female	[60,100)	496,690	6,336,498

Table A.2: Estimated # of smokers and non-smokers, per age/sex, used for calibration of wave 1 weights, waves 1–2 longitudinal weights, and waves 1–3 longitudinal weights.

Sex	Age	#smokers	#non-smokers
male	[18,25)	1,245,494	1,540,846
male	[25,35)	1,888,860	2,155,809
male	[35,50)	2,552,165	3,844,240
male	[50,60)	1,174,255	2,888,913
male	[60,100)	816,305	4,662,257
female	[18,25)	1,097,304	1,625,535
female	[25,35)	1,455,453	2,554,061
female	[35,50)	2,208,706	4,345,319
female	[50,60)	805,641	3,390,409
female	[60,100)	556,667	6,767,910

Table A.3: Estimated # of smokers and non-smokers, per age/sex, used for calibration of wave 2 cross-sectional weights and waves 2–3 longitudinal weights.

APPENDIX

Sex	Age	#smokers	#non-smokers
male	[18,25)	1,391,610	1,38,5500 ^a
male	[25,35)	2,134,533	1,702,484 ^a
male	[35,50)	2,800,673	3,676,370 ^a
male	[50,60)	1,209,176	2,825,445 ^a
male	[60,100)	800,228	5,403,087
female	[18,25)	1,199,683	1,512,691 ^b
female	[25,35)	1,620,603	2,277,889 ^b
female	[35,50)	2,444,563	4,165,938
female	[50,60)	1,114,971	3,137,396
female	[60,100)	646,106	7,501,500

Cells sharing the same letters (*a*, *b*) were collapsed when computing the wave 3 cross-sectional weights for newly recruited smokers.

Table A.4: Estimated # of smokers and non-smokers, per age/sex, used for calibration of wave 3 cross-sectional weights.

Région UDA	Départements
Région parisienne	Paris, Seine-et-Marne, Yvelines, Essonne, Hauts-de-Seine, Seine-Saint-Denis, Val-de-Marne, Val-d'Oise
Bassin parisien	Aisne, Ardennes, Aube, Calvados, Cher, Côte-d'Or, Eure, Eure-et-Loir, Indre, Indre-et-Loire, Loir-et-Cher, Loiret, Manche, Marne, Haute-Marne, Nièvre, Oise, Orne, Saône-et-Loire, Seine-Maritime, Somme, Yonne
Nord	Nord, Pas-de-Calais
Ouest	Charente, Charente-Maritime, Côtes-d'Armor, Finistère, Ile-et-Vilaine, Loire-Atlantique, Maine-et-Loire, Mayenne, Morbihan, Sarthe, Deux-Sèvres, Vendée, Vienne
Est	Doubs, Jura, Meurthe-et-Moselle, Meuse, Moselle, Bas-Rhin, Haut-Rhin, Haute-Saône, Vosges, Territoire de Belfort
Sud ouest	Ariège, Aveyron, Corrèze, Creuse, Dordogne, Haute-Garonne, Gers, Gironde, Landes Lot, Lot-et-Garonne, Pyrénées-Atlantiques, Hautes-Pyrénées, Tarn, Tarn-et-Garonne, Haute-Vienne
Centre est	Ain Allier, Ardèche, Cantal, Drôme, Isère, Loire, Haute-Loire, Puy-de-Dôme, Rhône, Savoie, Haute-Savoie
Méditerranée	Alpes-de-Haute-Provence, Hautes-Alpes, Alpes-Maritimes, Aude, Bouches-du-Rhône, Corse, Gard, Hérault, Lozère, Pyrénées-Orientales, Var Vaucluse

Table A.5: Grouping of the 100 Departments of France into the 8 UDA regions.

B Additional explanations for the $w_i^{(1)}$ weights

This section gives additional explanations on the derivation of the various multiplicative terms used in the computation of some of the $w_i^{(1)}$ weights. To this end, the following quantities must first be defined

- let $U_{\text{smk}} = \{1, \dots, N_{\text{smk}}\}$ be the population of smokers
- let $U_{\text{nsmk}} = \{1, \dots, N_{\text{nsmk}}\}$ be the population of non-smokers
- let S_{smk} be the sample of smokers; thus $S_{\text{smk}} \subseteq U_{\text{smk}}$
- let S_{nsmk} be the sample of non-smokers; thus $S_{\text{nsmk}} \subseteq U_{\text{nsmk}}$
- let $U^H = \{1, \dots, N_H\}$ be the population of households
- let S^H be the sample of households; thus $S^H \subseteq U^H$
- let U_{smk}^H , U_{nsmk}^H and U_{mixed}^H be, respectively, the population of households where all residents are smokers, the population of households where all residents are non-smokers, and the population of households where a mixed of smokers and non-smokers reside; thus $U^H = U_{\text{smk}}^H \cup U_{\text{nsmk}}^H \cup U_{\text{mixed}}^H$
- let $S_{\text{smk}}^H = S^H \cap U_{\text{smk}}^H$, $S_{\text{nsmk}}^H = S^H \cap U_{\text{nsmk}}^H$ and $S_{\text{mix}}^H = S^H \cap U_{\text{mix}}^H$
- let Q_{mix} be the event that the quota for household where a mixed of smokers and non-smokers reside is opened, and \bar{Q}_{mix} the event that the quota is closed
- let Q_{nsmk} be the event that the quota for household where a mixed of smokers and non-smokers reside is opened, and \bar{Q}_{nsmk} the event that the quota is closed
- let p_{mix}^H be the proportion of households where a mixed of smokers and non-smokers reside
- let p_{mix}^Q be the proportion of those mixed households contacted while the corresponding quota was opened
- let p_{nsmk}^H be the proportion of households where all residents are non-smokers
- let p_{nsmk}^Q be the proportion of those non-smoker households contacted while the corresponding quota was opened.

B.1 Computation of multiplicative terms in step 1 of section 2.2.1

In addition to the above definitions, let i be the i^{th} smoker from the population U_{smk} and let $h(i)$ be the household of the i^{th} smoker. Then,

$$\begin{aligned}
P(i \in S_{\text{smk}}) &= P(i \in S_{\text{smk}} \mid h(i) \in S_{\text{smk}}^H) \times P(h(i) \in S_{\text{smk}}^H) + \\
&\quad P(i \in S_{\text{smk}} \mid h(i) \in S_{\text{mix}}^H, Q_{\text{mix}}) \times P(h(i) \in S_{\text{mix}}^H, Q_{\text{mix}}) + \\
&\quad P(i \in S_{\text{smk}} \mid h(i) \in S_{\text{mix}}^H, \bar{Q}_{\text{mix}}) \times P(h(i) \in S_{\text{mix}}^H, \bar{Q}_{\text{mix}}) \\
&= P(i \in S_{\text{smk}} \mid h(i) \in S_{\text{smk}}^H) \times P(h(i) \in S_{\text{smk}}^H) + \\
&\quad P(i \in S_{\text{smk}} \mid h(i) \in S_{\text{mix}}^H, Q_{\text{mix}}) \times P(h(i) \in S_{\text{mix}}^H \mid Q_{\text{mix}}) \times P(Q_{\text{mix}}) + \\
&\quad P(i \in S_{\text{smk}} \mid h(i) \in S_{\text{mix}}^H, \bar{Q}_{\text{mix}}) \times P(h(i) \in S_{\text{mix}}^H \mid \bar{Q}_{\text{mix}}) \times P(\bar{Q}_{\text{mix}}) \\
&\approx P(i \in S_{\text{smk}} \mid h(i) \in S_{\text{smk}}^H) \times P(h(i) \in S_{\text{smk}}^H) + \\
&\quad P(i \in S_{\text{smk}} \mid h(i) \in S_{\text{mix}}^H, Q_{\text{mix}}) \times p_{\text{mix}}^H \times p_{\text{mix}}^Q + \\
&\quad P(i \in S_{\text{smk}} \mid h(i) \in S_{\text{mix}}^H, \bar{Q}_{\text{mix}}) \times p_{\text{mix}}^H \times p_{\text{mix}}^{\bar{Q}} \\
&\approx \frac{1}{N_S} \times \frac{1398}{15878 - 3355} + \frac{1}{N_A} \times \frac{784}{1162} \times \frac{784 + 1162}{15878 - 3355} + \\
&\quad \frac{1}{N_S} \times \left(1 - \frac{784}{1162}\right) \times \frac{784 + 1162}{15878 - 3355} \\
&= \frac{1}{N_S} \times 0.1621 + \frac{1}{N_A} \times 0.1048,
\end{aligned}$$

where N_S is the number of adult smokers residing in $h(i)$ and N_A is the number of adults residing in $h(i)$; estimates of p_{mix}^H and p_{mix}^Q were obtained from table A.6. Therefore,

$$w_i = \frac{1}{P(i \in S_{\text{smk}})} = \frac{N_S N_A}{N_A \times 0.1621 + N_S \times 0.1048},$$

which is the formula for $w_i^{(1)}$ in step 1 of section 2.2.1.

A similar procedure was used to compute the multiplicative terms in the construction of the wave 2 and 3 cross-sectional weights for newly recruited smokers, with the data from table A.7 replacing that of table A.6 for the wave 2 weights (see section 2.3.1), and the data from table A.8 replacing that of table A.6 for the wave 3 weights (see section 2.4.1).

B.2 Computation of multiplicative terms in step 1 of section 2.2.2

In addition to the definitions at the beginning of section B, let i be the i^{th} non-smoker from the population U_{nsmk} and let $h(i)$ be the household of the i^{th} non-smoker. Then,

$$\begin{aligned}
\text{P}(i \in S_{\text{nsmk}}) &= \text{P}(i \in S_{\text{nsmk}} \mid h(i) \in S_{\text{nsmk}}^H, Q_{\text{nsmk}}) \times \text{P}(h(i) \in S_{\text{nsmk}}^H, Q_{\text{nsmk}}) + \\
&\quad \text{P}(i \in S_{\text{nsmk}} \mid h(i) \in S_{\text{mix}}^H, Q_{\text{mix}}) \times \text{P}(h(i) \in S_{\text{mix}}^H, Q_{\text{mix}}) \\
&= \text{P}(i \in S_{\text{nsmk}} \mid h(i) \in S_{\text{nsmk}}^H, Q_{\text{nsmk}}) \times \text{P}(h(i) \in S_{\text{nsmk}}^H \mid Q_{\text{nsmk}}) \times \text{P}(Q_{\text{nsmk}}) \\
&\quad \text{P}(i \in S_{\text{nsmk}} \mid h(i) \in S_{\text{mix}}^H, Q_{\text{mix}}) \times \text{P}(h(i) \in S_{\text{mix}}^H \mid Q_{\text{mix}}) \times \text{P}(Q_{\text{mix}}) \\
&\approx \text{P}(i \in S_{\text{nsmk}} \mid h(i) \in S_{\text{nsmk}}^H, Q_{\text{nsmk}}) \times p_{\text{nsmk}}^H \times p_{\text{nsmk}}^Q \\
&\quad \text{P}(i \in S_{\text{nsmk}} \mid h(i) \in S_{\text{mix}}^H, Q_{\text{mix}}) \times p_{\text{mix}}^H \times p_{\text{mix}}^Q \\
&\approx \frac{1}{N_A} \times \frac{369 + 8810}{15878 - 3355} \times \frac{369}{8810} + \frac{1}{N_A} \times \frac{784 + 1162}{15878 - 3355} \times \frac{784}{1162} \\
&= \frac{1}{N_A} \times (0.0307 + 0.1048) \\
&= \frac{0.1355}{N_A},
\end{aligned}$$

where N_A is the number of adults residing in $h(i)$; estimates of p_{nsmk}^H , p_{mix}^H , p_{nsmk}^Q and p_{mix}^Q were obtained from table A.6. Therefore,

$$w_i = \frac{1}{\text{P}(i \in S_{\text{nsmk}})} = \frac{N_A}{0.1355} = N_A \times 7.380,$$

which is the formula for $w_i^{(1)}$ in step 1 of section 2.2.2.

A similar procedure was used to compute the multiplicative terms in the construction of the wave 2 and 3 cross-sectional weights for newly recruited non-smokers, with the data from table A.7 replacing that of table A.6 for the wave 2 weights (see section 2.3.2), and the data from table A.8 replacing that of table A.6 for the wave 3 weights (see section 2.4.2).

	FOYER NON FUMEUR QUOTA OUVERT		FOYER NON FUMEUR QUOTA ATTEINT		FOYER MIXTE QUOTA NON FUMEUR OUVERT		FOYER MIXTE QUOTA NON FUMEUR ATTEINT		FOYER FUMEUR		FOYER NON DETERMINE		TOTAL	
	NB	%	NB	%	NB	%	NB	%	NB	%	NB	%	NB	%
BASE UTILISEE	369	100	8,811	100	786	100	1,172	100	1,407	100	13,455	100	26,000	100
Hors Cible	0	0.0%	1	0.0%	0	0.0%	3	0.0%	8	0.1%	7,199	81.7%	7,211	27.7%
Ménages non francophones	0	0.0%	0	0.0%	2	0.0%	7	0.1%	1	0.0%	207	2.3%	217	0.8%
Ménages injoignables	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	2,694	30.6%	2,694	10.4%
MENAGES INTERROGEABLES	369	100	8,810	100	784	100	1,162	100	1,398	100	3,355	100	15,878	100
Résidences secondaires	0	0.0%	1	0.0%	1	0.0%	1	0.0%	0	0.0%	149	1.7%	152	1.0%
Ménages sans personne dans la tranche d'âge	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	16	0.2%	16	0.1%
Refus ménages immédiats	0	0.0%	1	0.0%	21	0.2%	36	0.4%	37	0.4%	2,780	31.6%	2,875	18.1%
Refus ménages différés	1	0.3%	0	0.0%	23	0.3%	61	0.7%	19	0.2%	0	0.0%	104	0.7%
Rendez-vous ménages non aboutis	0	0.0%	0	0.0%	0	0.0%	9	0.1%	3	0.0%	268	3.0%	280	1.8%
Inéligibles : foyers non-fumeurs quota atteint	0	0.0%	8,808	100.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	8,808	55.5%
Inéligibles : pas de fumeur de cigarettes	0	0.0%	0	0.0%	0	0.0%	0	0.0%	48	0.5%	0	0.0%	128	0.8%
Inéligibles : pas de fumeur de 100 cigarettes / vie	0	0.0%	0	0.0%	0	0.0%	37	0.4%	13	0.1%	0	0.0%	50	0.3%
Personnes impossibles	8	2.2%	0	0.0%	24	0.3%	65	0.7%	32	0.4%	142	1.6%	271	1.7%
Personnes injoignables	0	0.0%	0	0.0%	7	0.1%	33	0.4%	22	0.2%	0	0.0%	62	0.4%
PERSONNES INTERROGEABLES	360	100	0	100	708	100	840	100	1,224	100	0	100	3,132	100
Refus personnes sélectionnées	15	4.2%	0	0.0%	33	4.7%	35	4.2%	33	2.7%	0	0.0%	116	3.7%
Taux de refus de participer à l'étude		4.4%		0.0%		5.2%		5.3%		3.3%		31.6%		22.5%
PERSONNES INTERROGEEES	345	100	0	100	675	100	805	100	1,191	100	0	100	3,016	100
Fumeurs : refus de faire partie d'une cohorte	0	0.0%	0	0.0%	49	7.3%	141	17.5%	172	14.4%	0	0.0%	362	12.0%
Non-fumeurs : refus de faire partie d'une cohorte	49	14.2%	0	0.0%	38	5.6%	0	0.0%	0	0.0%	0	0.0%	87	2.9%
Rendez-vous questionnaires	0	0.0%	0	0.0%	6	0.9%	13	1.6%	25	2.1%	0	0.0%	44	1.5%
Abandons interviews	34	9.9%	0	0.0%	65	9.7%	77	9.7%	87	7.5%	0	0.0%	263	8.8%
Interviews réalisées	262	75.9%	0	0.0%	517	76.6%	574	71.3%	907	76.2%	0	0.0%	2,260	74.9%

Table A.6: Summary of wave 1 call-log.

APPENDIX

	FOYER NON FUMEUR QUOTA OUVERT		FOYER NON FUMEUR QUOTAPARTEINT		FOYER MIXTE QUOTA NON FUMEUR OUVERT		FOYER MIXTE QUOTA NON FUMEUR ATTEINT		FOYER FUMEUR		FOYER NON DETERMINE		TOTAL	
	NB	%	NB	%	NB	%	NB	%	NB	%	NB	%	NB	%
BASE UTILISEE	109	100	2,254	100	91	100	423	100	409	100	3,619	100	6,905	100
Hors Cible							11	2.6%	6	1.5%	831	23.0%	848	12.3%
Ménages non francophones											48	1.3%	48	0.7%
Ménages injoignables	1	0.9%					1	0.2%			915	25.3%	917	13.3%
Résidences secondaires			30	1.3%							70	1.9%	100	1.4%
Ménages sans personne dans la tranche d'âge											1	0.0%	1	0.0%
Personnes impossibles	1	0.9%					9	2.1%	4	1.0%			14	0.2%
MENAGES INTERROGEABLES	107	100	2,224	100	91	100	402	100	399	100	1,754	100	4,977	100
Refus ménages avant de parler à la personne sélectionnée	14	13.1%	7	0.3%	10	11.0%	59	14.7%	37	9.3%	1710	97.5%	1,837	36.9%
Rendez-vous ménages non aboutis							16	4.0%	4	1.0%	44	2.5%	64	1.3%
Inéligibles : foyers non-fumeurs quota atteint			2216	99.6%	1	1.1%	3	0.7%					2,220	44.6%
Inéligibles : pas de fumeur de cigarettes														
Inéligibles : pas de fumeur de 100 cigarettes / vie														
Personnes injoignables	1	0.9%			1	1.1%	12	3.0%	6	1.5%			20	0.4%
PERSONNES INTERROGEABLES	92	100	1	100.0%	79	100	299	100	345	100	100	100	816	100
Refus personnes sélectionnées	4	4.3%	1	100.0%	4	5.1%	39	13.0%	19	5.5%			67	8.2%
Taux de refus de participer à l'étude		16.8%		0.4%		15.4%		24.4%		14.0%		97.5%		38.3%
PERSONNES INTERROGEEES	88	100			75	100	260	100	326	100			749	100
Fumeurs : refus de faire partie d'une cohorte					4	5.3%	35	13.5%	42	12.9%			81	10.8%
Non-fumeurs : refus de faire partie d'une cohorte	20	22.7%			3	4.0%							23	3.1%
Rendez-vous questionnaires					1	1.3%	30	11.5%	19	5.8%			50	6.7%
Abandons interviews	2	2.3%			2	2.7%	9	3.5%	8	2.5%			21	2.8%
Interviews réalisées	66	75.0%			65	86.7%	186	71.5%	257	78.8%			574	76.6%

Table A.7: Summary of wave 2 call-log.

	FOYER NON FUMEUR QUOTA OUVERT		FOYER NON FUMEUR QUOTA ATTEINT		FOYER MIXTE QUOTA NON FUMEUR OUVERT		FOYER MIXTE QUOTA NON FUMEUR ATTEINT		FOYER FUMEUR		FOYER NON DETERMINE		TOTAL	
	NB	%	NB	%	NB	%	NB	%	NB	%	NB	%	NB	%
BASE UTILISEE	96	0.60%	2,597	16.33%	53	0.33%	584	3.67%	1,190	7.48%	11,386	71.58%	15,906	100
Hors Cible	6	0.3%	0	0.0%	2	0.1%	14	0.6%	72	3.0%	2,292	96.1%	2,386	100.0%
Ménages non francophones	2	1.0%	0	0.0%	1	0.5%	5	2.5%	3	1.5%	192	94.6%	203	100.0%
Ménages inéligibles	1	0.0%	0	0.0%	3	0.1%	35	1.2%	25	0.9%	2,875	97.8%	2,939	100.0%
Résidences secondaires	0	0.0%	0	0.0%	0	0.0%	4	0.4%	88	6.3%	1,016	93.4%	1,088	100.0%
Ménages sans personne dans la tranche d'âge	0	0.0%	0	0.0%	0	0.0%	0	0.0%	6	25.1%	17	73.9%	23	100.0%
Personnes impossibles	2	1.0%	0	0.0%	4	1.9%	46	22.0%	25	12.0%	132	63.2%	209	100.0%
MENAGES INTERROGABLES	85		2,597		43		480		991		4,862		9,058	100%
Refus ménages avant de parler à la personne sélectionnée	15	0.3%	0	0.0%	8	0.2%	88	1.8%	356	7.3%	4,415	90.4%	4,882	100.0%
Rendez-vous ménages non aboutis	0	0.0%	0	0.0%	0	0.0%	19	9.6%	12	6.1%	167	84.3%	198	100.0%
Inéligibles : foyers non-fumeurs quota atteint	0	0.0%	2,581	86.3%	0	0.0%	9	0.3%	309	10.3%	90	3.0%	2,989	100.0%
Inéligibles : pas de fumeur de cigarettes	0	0.0%	0	0.0%	0	0.0%	0	0.0%	1	100.0%	0	0.0%	1	100.0%
Inéligibles : pas de fumeur de 100 cigarettes / vie	0	0.0%	0	0.0%	0	0.0%	4	100.0%	0	0.0%	0	0.0%	4	100.0%
Inéligibles : pas de fumeur mensuel	0	0.0%	0	0.0%	1	4.0%	8	32.0%	3	12.0%	13	52.0%	25	100.0%
Personnes inéligibles	0	0.0%	0	0.0%	1	4.0%	8	32.0%	3	12.0%	13	52.0%	25	100.0%
PERSONNES INTERROGABLES														
Refus personnes sélectionnées	0		7		2		34		23		105		171	
Taux de refus de participer à l'étude														
PERSONNES INTERROGÉES														
Fumeurs : refus de faire partie d'une cohorte	0		0		0		21		11		32		64	
Non-fumeurs : refus de faire partie d'une cohorte	0		0		1		8		3		13		25	
Rendez-vous questionnaires	0		9		0		30		18		26		83	
Abandons interviews	0		0		0		259		255		1		616	
Interviews réalisées	70		0		31		259		255		1		616	

Table A.8: Summary of wave 3 call-log.