In fighting smoking, would later not be better than sooner?

May 2014

Michel Grignon¹
Thomas Renaud² ³

¹ Economics and CHEPA, McMaster University, Hamilton, Ontario, Canada
² Independent Consultant, University Paris Dauphine, France
³ Disclaimer: This study was funded by Pfizer Europe.
CHEPA WORKING PAPER SERIES

The Centre for Health Economics and policy Analysis (CHEPA) Working Paper Series provides for the circulation on a pre-publication basis of research conducted by CHEPA faculty, staff and students. The Working Paper Series is intended to stimulate discussion on analytical, methodological, quantitative, and policy issues in health economics and health policy analysis. The views expressed in the papers are the views of the author(s) and do not necessarily reflect the views of the Centre or its sponsors. Readers of Working Papers are encouraged to contact the author(s) with comments, criticisms, and suggestions.

NOT FOR CITATION WITHOUT PERMISSION
1 Introduction and short review of evidence

1.1 The question and our objectives

In this study, we address the following question: Where are resources to prevent smoking and reduce smoking prevalence (the proportion of the adult population who is smoking on a regular basis) most efficiently put to use, in preventing youth from ever starting smoking or in helping seasoned smokers quit? We use mortality attributable to tobacco (MAT) as our outcome measure in deciding what is efficient: our criterion, therefore, is cost per life saved (death averted). This means that we take an instrumental approach to smoking: it is not smoking *per se* that is bad, but smoking in that it causes premature mortality $^1$.

The question can be re-framed as follows:

- Policy makers have a choice to make: where should the next Euro (respectively Pound, Dollar) allocated to smoking reduction be spent, in preventing youth initiation or in helping smokers quit? It is often argued that both avenues should be pursued at the same time, but this does not address the issue, since, at one point, one has to make a decision relative to the next unit of resource to spend.

- Preventing one more youth to smoke will have an effect in 40 years, and depends crucially on what youth is prevented from starting smoking: if youth susceptible to smoking prevention campaigns are those who would have quit anyway before age 30, the effect of campaigns targeted toward the young on MAT will be minimal and could even be 0. In any case, it will take a long time to be effective.

- Helping smokers to quit, especially smokers at risk (i.e. smokers with more than 20 years of smoking) will have an immediate effect on MAT. Of course, there is also a chance that smokers able to quit after more than 20 years of smoking are precisely those who would have quit anyway (with or without help); in that sense, helping smokers to quit would only speed up the process for those already willing to quit rather than save those hard-core smokers not willing to quit in the first place. However, it can be argued that, in this very specific sub-population with a very high risk of MAT (smokers with more than 20 years of smoking), any year cut from smoking improves vital prognosis substantially and will have a significant effect on mortality.

---

$^1$A word of caution is warranted here: smoking certainly has an impact on health beyond premature mortality; those living with throat or larynx cancer or Chronic Obstructive Pulmonary Disease (COPD) experience dramatically reduced quality of life. This study, restricted as it is to premature mortality due to smoking (and even more restricted to premature mortality due to lung cancer as shown below), can be seen as a first step toward a more comprehensive understanding of the impact of smoking prevention and cessation tools on health-related outcomes of smoking. Our choice to restrict ourselves to MAT is based solely on epidemiological data availability, not on any judgement on the relative "value" of premature death versus prolonged life in poor quality. However, we remain clearly "instrumental" in the sense that being able to smoke without any adverse health effect would be fine from our perspective.
• Of course, as is the case with any such question of marginal efficiency, it all depends on where we stand as of now: if it is the case that youth prevention is effective on future marginally attached smokers only and that quitting aid helps many smokers in the high risk population, then next Euros should go to quitting programmes for a while; however, after so many long-time smokers have quit, it may be the case that quitting aid programmes will become ineffective (either because they will not work on hard-core smokers or because they will cost too much per hard-core smoker to quit) and it might be time to go back to preventing smoking among teenagers. As any economic evaluation, this one is "here and now."

• Last, we do not use information on the cost per unit of outcome (youth prevented from smoking or smoker helped to quit) and are content with estimating the effect of both avenues to reducing smoking prevalence (preventing initiation versus helping quits) on MAT. It goes without saying that a comprehensive economic evaluation would have to take account of such information. A cost-benefit analysis would also have to take into account the welfare effects of preventing smoking: if a teenager who would have become an easy quitter at age 25 is prevented from smoking, they will lose 10 years of "pleasure" derived from smoking that does not inflict any substantial health risk; since they are willing to pay for the cigarettes they smoke, standard economics assume there is a value to that pleasure, and that depriving one of it will have a detrimental effect on societal well-being.

In order to answer this question of marginal efficiency we need to know more about how smoking affects its users (in harsher words, how it kills them): a key feature of the answer lies in the effect of duration of smoking on MAT, relative to the effect of ever smoking. If it is the case that ten years as a smoker do not really affect longevity but that 30 years do, then there is a case for spending our Euros on cessation. This means we have to decompose MAT into the effect of duration and the effect of being a smoker today. The way MAT is calculated as of now (in France, but also worldwide, with the very recent exception of a series of sophisticated micro-simulations conducted in the US that we briefly describe below) does not allow us to separate the effect of changes in the rate of initiation from that of an increase in cessation; in other words, given how MAT is calculated today, a decrease of 1,000 in the number of smokers aged 20 (this age is taken as representative of younger smokers with mechanically short durations of smoking) is supposed to have the same effect on MAT as a decrease of 1,000 in the number of smokers aged 50 who have continuously smoked for the past 30 years. The standard method, as described by Hill and Laplanche (2003, 2004), works as follows (it is derived from a method put forward by the WHO, described in Appendix 0):

• The number of deaths caused by lung cancer attributable to tobacco is estimated as the total number of deaths caused by lung cancer minus the number of these deaths among non smokers. The latter is derived
from rates (per 100,000) estimated in the Cancer Prevention Study II (CPSII), conducted in the US, by age and sex, applied to the French population: if $R_{x,i}^{NS}$ is the lung cancer mortality ratio among non-smokers at age $x$, for sex $i$ (observed in CPSII), the number of lung cancer deaths attributable to tobacco in France is calculated as follows: $MLCAT = \sum_{x=1}^{x=100} LCD^{x,i} - RR_{x,i}^{NS}$ POPF$^{x,i}$ LCD the number of deaths caused by lung cancer in France for age $x$ and sex $i$, and POPF the number of individuals of age $x$ and sex $i$ in the French population.

- For causes of death other than lung cancer, a relative risk (smokers relative to never smokers) of dying for each disease, RR (estimated on the CPSII study) is applied to an observed prevalence rate of the disease in the population, $P$. The surplus mortality due to the risk factor (number of deaths that happened that would not have taken place without the risk factor, i.e., if the risk factor were 0) is $P.(RR-1)$, and baseline risk of dying is $P.RR + (1-P) = P.(RR-1) + 1$ (mortality risk being 1 among those without the risk factor). Therefore, the proportion of total deaths that can be linked back to the risk factor is: $PAF = \frac{P.(RR-1)}{P.(RR-1)+1}$ PAF is 0 when either $P=0$ (no prevalence of the risk factor) or $RR=1$ (no risk associated with the factor beyond baseline risk) and tends toward 1 when $P=1$ AND RR is very large (toward infinity). Multiplying the number of cause-specific deaths by the PAF for that cause of death provides MAT for that cause. Adding over all causes of death yields total MAT.

Based on this description of how MAT is calculated today, it is clear that these methods do not tell us anything on the differential effect on mortality of changes in the prevalence of smoking due to prevention (lower initiation rates) and treatment (higher quitting rates).

In this study, we propose a novel method (based on simulations) to estimate MAT as a result of smoking duration (for smokers and ex-smokers), intensity (number of cigarettes smoked per day) of smoking while a smoker, and duration since cessation for ex-smokers. This will allow us to simulate the effect on mortality of various ways of reducing prevalence, either through decreasing initiation or through increasing cessation. Our framework relies on several key empirical assumptions that we review in turn in the next sub-section. Before, we outline what the study does:

- Our study is based on the idea that MAT is a function of duration of smoking, as well as of duration in the state of ex-smoker (cessation). In other words, it is not so much smoking prevalence in itself that matters when one is interested in MAT, but rather the distribution in the population of duration in the smoker and ex-smoker states.
- However, most of the focus so far has been on aggregate smoking prevalence (public policy seems to be targeted on reducing prevalence, without too much questions on how it reduces the proportion of long-time smokers).
• If anything, there is more interest among policy makers (and some epidemiologists) in reducing youth prevalence than among older adults (40 years old and older), even though these older adults are more likely to be long-time smokers. We posit that such a focus is not supported by any evidence (at least any evidence on its effect on MAT) and may be the cause of a waste in resources in the fight against MAT.

• What we will do:

1. Present evidence on the distribution of duration of smoking (number of years in the "smoker" state) by sex and for various birth cohorts, to document how the recent decrease in overall prevalence in France took place: how much of it is the result of fewer youth entering smoking versus fewer smokers continuing smoking beyond age 30? It is important to note that this will be the first time such evidence will be made available in the case of France and, as far as we know, for any European country. The only comparable set of evidence available is for the US and indicates a substantial reduction in the average duration of smoking; Using surveys conducted in the US, Mannino et al. (2000) show that smoking prevalence at age 34.5 is substantially lower (at 47%) among white males born in 1951-55 than among white males born in 1936-40 (at 67%), and even than for white males born in 1901-05 (57%). The same trend is observed among (white) female smokers, with prevalence at age 34.5 having increased from 18% for the birth cohort 1901-05 (this was observed in 1936-40) to 52% for the birth cohort 1936-40 (observed in 1971-75) before declining to 39% for the most recent cohort (born in 1951-55, therefore observed in 1986-90). Similarly, Pierce and Gilpin (1996) show a clear linear pattern in the decline in age at which smoking prevalence in a cohort is at half the peak value (usually reached at around age 25): the slope is -2.33 for (white) males and -2.54 for (white) females, and all observations are on the regression line. If the relationship remains stable beyond their youngest cohort (born 1930-34), we should expect to see half of smokers having quit before age 35 for the cohort born 1975-79 (and even, extrapolating even further, before age 25 for the cohort born 1990-95). Anderson et al. (2012) confirm the recent increase in cessation rates at each age (for cohorts born after 1965) in the US.

2. Run simulations of the effect on MAT of reducing prevalence through preventing initiation (entering smoking) versus accelerating quits (helping older smokers quit). These simulations will be based on epidemiological evidence on the effect of duration of smoking and duration of cessation on mortality. It will also be based on a set of assumptions that are discussed below

(a) Nobody starts smoking after the age of 30 and the distribution of starting ages is highly concentrated around 18 (Burns,
2000; Mannino et al., 2000 and Anderson et al., 2012 for the US; Grignon, 2009, for France). Empirical evidence presented in the first part of the study will show that this is a reasonable assumption for all cohorts except one (women born in years 1935-45 increased their smoking prevalence in the 1970s, when they were in their 30s, certainly due to a period effect allowing women to smoke whereas they had been prevented before by social norms).

(b) Duration is the most important factor of MAT. More specifically, starting age does not matter per se (in the reasonable range between 15 and 30) and there is no gain in postponing initiation from a mortality point of view (it is still disputed whether starting later will make quitting easier or whether late starters tend to also quit later); second, the number of cigarettes smoked per day certainly has an impact, but it is of a second order compared to the number of years as a smoker. There are three reasons for that: first, the relative risk of MAT seems to be less sensitive to daily dose (thereafter: intensity) than to duration of exposure in epidemiological studies (even though this finding is not universal, see Rachet et al. 2004); second the distribution of intensity is more highly concentrated (between 15 and 25 cigarettes per day) than the distribution of number of years as a smoker (between 5 and 50 per life); third, tobacco control policies seem to have much more impact on reducing prevalence (increasing quits or preventing initiations) than on reducing intensity.  

(c) There is something called cessation (or quit) and it is forever. Obviously, many smokers attempt to quit, then relapse and try again. The literature tends to consider these smokeless spells as being part of the smoking "career", even though they can last a few years for some smokers.

This will be one of the first time such simulation-based evidence is provided. 3 Most statements on MAT tend to be based on a very rough linkage between aggregate prevalence and total number of deaths, as if the ratio of MAT per smokers were constant (independent of the distribution of duration of smoking and cessation). As a result, what is published on MAT (a now well-publicized figure of 60,000 deaths due to smoking in France) might be wrong, and, more importantly, it does not provide policy-makers and the general public the right pieces of information to evaluate the effectiveness and efficiency of policy interventions to reduce smoking.

---

2 The latter is linked to the former: it is because smokers find it almost impossible to gradually decrease intensity, and easier to quit cold turkey, that intensity is highly concentrated around 20.

3 See, though, the recently published special issue of Risk Analysis on the effects of tobacco control on mortality using complex microsimulations to address the same question in the case of the US.
1.2 Brief review of empirical evidence supporting our framework

1.2.1 Effect of duration, cessation, and intensity on MAT

The literature on the effect of smoking on mortality and morbidity is vast but, here, we focus on the small subset of studies calculating relative risks in relation to duration and intensity of smoking.

The bulk of the evidence on mortality and morbidity as a function of duration and intensity of smoking is on lung cancer. A few studies show a benefit of quitting smoking (reduced mortality) even among patients with lung cancer (Ebbert et al., 2005) or after a myocardial infarction (Wilson et al., 2000), but most studies are about the link between smoking and incidence of or mortality due to lung cancer. The main finding from the epidemiology of smoking and lung cancer is that duration of smoking matters a lot, more so than intensity (number of cigarettes smoked per day). It also shows surprising differences by sex.

A retrospective study of deaths due to lung cancer in the US (Mannino et al., 2000) shows a strong effect of duration on mortality among cohorts born between 1901 and 1951 but the authors cannot compare it to the effect of daily dose. Similarly, Knoke, Burns, and Thune (2008), using the first wave of the CPS study (CPS I), show a strong effect of cessation on mortality from lung cancer, the risk being reduced by half after 5 years if the smoker quits at 40 (and 10 years if they quit later) and by 90% after 15 to 20 years. Their study controls for the quitting ill effect, the fact that some smokers quit when they learn they have lung cancer.

The seminal paper on the effect of smoking (on incidence of lung cancer rather than mortality, but, since lung cancer is rarely treatable, both concepts are closely related) was published in 1978 by Richard Doll and Richard Peto (Doll & Peto, 1978). The underlying biological model is that carcinoma develops in stages and, as a result, the mortality response to smoking will not be linear in duration (if the response is linear at each stage, the overall response will be quadratic). Using a cohort of male British doctors followed between 1951 and 1971, they adjust a model based on the assumption that incidence of lung cancer is a function of intensity (daily number of cigarettes per day) and duration (number of years of smoking). Using a Maximum Likelihood estimator, they adjust the incidence of lung cancer to a multiplicative model of intensity and duration, assuming away any interaction effect on relative risk \(^4\):

\[
I = 0.273 \times 10^{-12} \times (\text{cigarettes/day} + 6)^2 \times (\text{age} - 22.5)^{1.5}.
\]

(The ML estimator finds the exponent and the constant, whereas the specific parameters (6 and 22.5) are entered manually and several MLEs are run for several pairs of these \textit{ad hoc} parameters). The empirical values estimated for the parameters suggest that doubling intensity from 20 to 40 cigarettes per day increases incidence by a factor 2.5 (and reducing it from 20 to 10 reduces incidence by 60%) but adding 5 years of smoking from 30 to 35 increases incidence by a factor 10 (and adding 5

\(^{4}\text{They test for such interaction effects and conclude they are non significant.}\)
years at age 35 by a factor 4.6]. Overall, still being a smoker at age 50 increases incidence relative to quitting at age 30 by an almost non credibly large factor 346. Duration clearly was, in that population of British doctors in the middle of the twentieth century, the major driver of lung cancer, far ahead of intensity. It should be made clear here that any daily intensity below 10 will not really affect MAT, and can be considered for our purpose similar to not being a smoker at all.

A follow up over 14 years (1980-93) of approximately 50,000 women aged 50 and older and registered with a breast cancer monitoring registry in Canada confirms Doll and Peto’s finding that duration matters more than intensity: controlling for potential confounding factors such as physical activity and BMI (as well as education). Zhang et al. (2005) find that the hazard ratio per year of smoking is double that for cigarettes per day (1.08 versus 1.04). They show that, compared to current smokers aged 50 and older, those who quit more than 20 years ago have a relative risk of 0.04, and, after 0 to 9 years without smoking, the relative risk has gone down by 60%. Quitting before 30 means that the relative risk of dying from lung cancer is exactly the same as that of never smokers and quitting before age 39 increases the risk by 1.5 only (recall that the baseline risk is very low).

The Cancer Prevention Study II (CPSII) conducted in the US between 1982 and 1988 on 600,000 individuals of both sexes also confirms Doll and Peto (1978) as well as Zhang (2005) findings that duration matters more than intensity (Flanders et al., 2003). They estimate separate relationships for 10 year age categories (40 to 49, 50 to 59, 60 to 69, and 70 to 79) and sex, between mortality due to lung cancer on one hand and intensity and duration on the other hand, all relationships following the same functional specification as: the relationships being estimated separately for various age categories and the two sexes: \( \hat{M}_s = e^{\alpha_\gamma . D \beta . t \gamma^\gamma} \), \( x \) indexing age category, and \( s \) sex, \( D \) being duration and \( I \) intensity. We provide the detailed estimates for the three parameters \( \alpha, \beta, \gamma \) in section 4.1.2.

The parameter estimated by Flanders et al. (2003) indicate that, among male smokers, doubling intensity from 20 to 40 cigarettes per day at age 40 (at duration 20, to be precise) doubles the relative risk but that increasing duration from 20 to 30 years multiplies risk by between 9 and 14 (depending on intensity; the higher the intensity, the lower the effect of duration). The effect of daily dose is similar among women, but the effect of duration is much smaller, the risk being multiplied by a factor between 4 and 8 depending on daily dose. Even for female smokers, though, duration matters much more than daily dose.

The only dissenting voice is Rachet et al. (2004), in a study of approximately 2,000 males aged 35 to 70 living in the Montreal area followed from 1979 to 1985: they find a strong intensity effect until 30 cigarettes per day, the odds ratio of lung cancer doubling for each 10 more cigarettes per day. They also find a strong effect of duration, with odds ratios increasing by 3% for each year as a smoker. Their findings are hard to compare to Doll and Peto’s because they use a shorter follow up period (six years instead of twenty) and a population which seems to smoke much more than male doctors in the UK. Also, findings
are presented as odds ratios instead of absolute incidences. They also find that starting age does not seem to matter on the probability to develop lung cancer (once duration is controlled), and that cessation works as follows: the absolute incidence remains constant once the smoker has quit. Moreover, for those who quit at age 35, absolute incidence is as low as that for never smokers (relative risk is between 5 and 2 according to duration since cessation but it applies to a baseline close to 0). Last, after ten years without smoking, differences in odds ratios by intensity smoked disappear.

Beyond lung cancer, one meta-analysis finds an effect of smoking on the incidence of and mortality from colorectal cancer, with a strong effect of duration of smoking (Liang et al., 2009) and another one on incidence of COPD, chronic bronchitis, and emphysema (Forey et al., 2011), finds higher risks among current smokers but no clear effect of duration of smoking (in this case, intensity seems to matter).

Streppel et al. (2007) report the result of a longitudinal study of smoking behaviours and mortality conducted on men born 1900-20 an living in Zutphen (the Netherlands) in 1960. They were followed between 1960 and 2000. In 1960, when they were on average 49 years of age, 89% of these men smoked daily (and another 6% had quit less than 10 years before, therefore the incredibly high rate of 95% of males aged 40 in Zutphen were smokers in the year 1950). Controlling for competing risks (food and alcohol consumption as well as BMI), they find that one more year of smoking increases the hazard of all cause mortality by 1.2% and each cigarette smoked within a day increases it by 1.1%. Duration is the only driver of mortality for cardiovascular diseases, chronic obstructive pulmonary disease, and lung cancer (daily dose is not significantly associated with the hazard ratio for these two causes of death). Intensity is associated with a higher hazard ratio than duration for all cancer and coronary heart diseases but, in both cases, ratios are not significantly different from 0. Their results show that smoking until 40 and quitting at that age means a loss of 1.9 years of life compared to a never smoker but a gain of 1.3 years compared to quitting at 50 and 1.8 years compared to quitting at 60. It can also be seen in their findings that the relationship between the number of cigarettes smoked per day and mortality is not robust: the effect is quite strong from one to cigarettes per day, which seems to indicate that other variables are not fully controlled in the model.

The evidence on the effects of cessation for the US population is summarized in the report of the Surgeon General for 1990 (reference: Department of Health and Human Services. The Health Benefits of Smoking Cessation. Washington, DC: US Government Printing Office, 1990. (DHHSPublication No. CDC90-8416): for light smokers (20 a day and less), all-cause mortality of ex-smokers is the same as for never-smokers after 15 years cessation; for heavy smokers (21 a day or more), the mortality risk relative to non-smokers is between 1.10 and 1.40 after 16 years without smoking.
1.2.2 Is the effect of duration constant over time?

Our simulation relies on the idea that duration (and, to a lesser extent, intensity) is the main cause of mortality (and incidence of main causes of death) and, moreover, that the causal relationship, being biological, remains stable over time. Of course, if it turns out that individuals with the same level of exposure to smoking (same duration and intensity) are more (less) likely to die from it or develop cancer now than twenty years ago, the goal of controlling smoking-related damage (on mortality) through helping smokers to quit might have to be reconsidered. Lethality might change for the following reasons: cigarettes are less (more) poisonous today than in the past, or smokers select differently than in the past. The former seems unlikely, since no study so far has been able to show that filtered cigarettes or low tar cigarettes were actually less hazardous than cigarettes smoked in the past. The latter is possible: if only really healthy individuals take the chance of smoking, the habit might become less lethal; on the other hand, if more vulnerable individuals are more likely to smoke, smoking might have become more lethal for the same level of exposure. This would be a case of confounding factors, such as education or occupation, or even perhaps, of unobservable heterogeneity.

One study (Mehta and Preston, 2012) seems to suggest that this is precisely the case: using surveys conducted on the general population in the US they find that the relative mortality of smokers (compared to non-smokers) after age 60 has increased by 25% between 1987 and 2006, holding smoking duration constant, and they suggest that "Smokers may have become more adversely selected on other health-related variables." However, a careful reading of their study shows that they do not actually control for the duration of smoking across cohorts (which would be the duration of smoking of current and former smokers as a whole), but separately for current and former smokers. Duration among current smokers is constant over time (which signals that starting age is itself stable) but it is decreasing among former smokers (they quit earlier). Mehta and Preston do not present their findings on the time trend among former smokers and find a positive time trend for current smokers (smaller than when duration is not added to the model, but still significantly positive). This seems to indicate that smokers who do not quit select differently now than they did twenty years ago; contrary to the "quitting ill effect" mentioned in Knoke, Burns and Thune (2008), which states that those who quit are usually sicker than those who do not quit, it seems here that those who quit are healthier on average. This is confirmed by a table showing more physical activity and less mental health problems among former smokers compared to current smokers. The issue raised here is a real one: if quitters are healthier on average, encouraging smokers to quit en masse may not be enough if residual smokers are the sickest of all. This interpretation (that the effect found by Mehta and Preston, 2012 is an artifact due to a selection effect of quitters versus current smokers) is supported by two retrospective studies of mortality by cohort (also in the US):

- Mannino et al. (2000) show that lung cancer mortality at a given duration is stable across birth cohorts (page 97). It should be kept in mind, though,
that, their study being retrospective, they stop at birth cohort 1951-55.

- Burns, Anderson, and Gray (2011) find that lung cancer mortality by cohort in the US is well predicted by a model applying the effect of duration on mortality (derived from CPSII) to variations in duration in smoking of various cohorts.

Note though that these two studies focus on lung cancer mortality whereas Mehta and Preston (2005) is about all cause mortality (reinforcing the selection interpretation of their findings).

Our main conclusion from this short review of the literature is that the relationship between duration and mortality is stable over time but that it is worth allowing for selection effects, if healthier individuals quit sooner. The only study testing this selection effect on individual data (Lahiri and Song, 2000, on the US Health and Retirement Survey) rather than time series indeed finds a selection effect but in the opposite direction as suggested by Mehta and Preston (2012) on time series: quitters are more likely to have received signals that smoking affected their health and current smokers are a more robust lot, conditional on age.

2 Data and Methods

2.1 General methodological framework

As a reminder, our objectives are as follows:

1. To provide the best available evidence on smoking prevalence by age and cohort, so that we are able to estimate the distribution by duration of smoking of the stock of smokers in any given year, as well as the distribution by duration since cessation of the stock of ex-smokers in any given year.

2. To interpret changes in the age profile of smoking across cohorts: is the drop in smoking prevalence observed in the early 2000s the result of fewer youth starting smoking or more middle-aged smokers quitting (meaning, smokers smoking fewer years on average)?

3. To estimate current mortality attributable to tobacco (MAT) based on refined evidence on smoking by duration and time since cessation (instead of the crude estimates used in the literature to date, which infer deaths from overall prevalence without allowing for the effect of duration and time since cessation).

4. To estimate MAT in "what if" scenarios of smoking prevalence: what would be the effect of a given drop in prevalence if it were channelled mostly through a decline in initiation (fewer youths starting smoking) versus more smokers quitting at a younger age?
2.2 Data

We use a pseudo-panel approach to reconstitute prevalence data by age for successive cohorts. In a pseudo-panel method, we pool cross-sectional surveys and calculate prevalence rates by age within each cross-section. We then re-arrange these rates to produce profiles by age for various cohorts: if we know $S_{30-34,T}$, the prevalence for age [30-34] in year T, and $S_{35-39,T+5}$, the prevalence for age [35-39] in year $T+5$, we are in a position to reconstitute a portion of the age profile (ages 30 to 39) for a specific cohort (born in T-30:34). The more cross-sections we pool, the longer the age profiles for the more cohorts we can reconstitute.

Such an exercise relies on accessing surveys measuring smoking prevalence in a consistent manner over a long period of time. More specifically, we need the following:

- a consistent definition of "smoker" across surveys (some surveys ask about regular consumption whereas others are more lenient and include occasional smokers);
- data collected at regular and constant time intervals, say every 5 years;
- age categories defined in a consistent way in publications (or the ability to reconstruct these categories if age is continuous in the survey);
- publications of prevalence rates by sex in all surveys, since smoking histories vary in a fundamental way across genders.

The last three conditions are particularly important for us: as we detail below, we can use sensitivity analysis to allow for "not-as-consistent-as-hoped-for" definitions of smoker across surveys.

Three institutions have collected, analysed, and published such data on smoking prevalence for several years in France: INPES (National Institute for Health Education and Promotion, Baromètre santé since 1992, previously known as "Enquêtes CFES"), INSEE (National Institute of Statistics, Enquête Santé), and IRDES (Institute for Research in Health Economics, Enquête Santé et Protection Sociale). In this study we use the Baromètre Santé (and enquêtes CFES prior to 1992 \(5\)) only since its cross-sectional surveys go back the longest in time with a reasonable gap between each survey (IRDES data start in the early 1990s and INSEE runs its survey every 10 years or so, with an erratic frequency).

Baromètre santé is a general population survey conducted since 1977 with consistent questions on smoking prevalence, even though there are variations across years in sample size and sampling procedure (for the better in recent years). Characteristics of the surveys are summarized in Table 1: year of production, general design, population selection method, tobacco consumption measurement and sample size.

Table 1 [here]

---

\(^5\) From now on, we use Baromètre Santé as a generic name for these surveys
As can be seen from table 1:

- Smoking status consistently includes regular as well as occasional smokers for all years since 1977.

- However, age categories and the ranges of ages included in the survey vary across years:
  - Age range: minimum age went from 18 to 15 to 12 and back to 15, and maximum was extended from 75 to 85. This is not so much of a problem for us since our focus is really on the 20 to 60 year olds.
  - More problematic is the lack of consistency in the breakdown into age categories. We go back to the issue and how we addressed it below.

- Sampling was greatly improved since the first versions (moving from quotas to random) and the sample size massively enlarged (from 1,000 persons in 1977 to almost 30,000 starting 2005). With 1,000 only, breakdowns by age and sex are less robust than could be hoped for.

Pooling these data, we can create a roughly consistent collection of prevalence rates for tobacco consumption by sex and age groups over a span of 35 years, from 1975 to 2010. We acknowledge the very generous help of François Beck and Romain Guignard (INPES) who allowed us to access individual data for recent years and create our own age categories based on a continuous age variable as collected in the surveys. However, for older surveys (before 1992), raw data are not accessible and we had to make do with age categories as published.

2.3 Extrapolations on data and alternative scenarios

We now present the steps we followed to reconstitute cohort-based prevalence data and overcome inconsistent age categories across years in the survey.

2.3.1 Interval between data measurements

Our first problem is that surveys were not conducted exactly five years apart over the whole period. We have to use the survey conducted in 1992 to approximate prevalences by age and sex in 1990, the 1986 survey for prevalences in 1985, the 1981 survey for prevalences in 1980 \(^6\) and the 1977 survey for prevalences in 1975.

\(^6\) We use three additional surveys not presented in table 1, also conducted by INPES (CPES at the time), around 1980 (January 1979, November 1979, and during the year 1982). Aggregating and smoothing results from 4 surveys (1981 plus the three aforementioned) increases the precision of the prevalence rates estimated for 1980, which is welcome given the limited sample size of the 1981 survey. This alternate data set will be considered in section 5.
2.3.2 Differences in age categories over years

Discrepancies in age categories across surveys are presented in table 2 below.

| Table 2 here |

As can be seen from table 2, we are faced with the following four problems:

1. No respondents over 75 except in 2010.

2. Respondents aged 15 to 17 are not included in surveys conducted in 1977, 1981, 1992 and 1995. This is of course problematic since we know that the average age at initiation is below 18 in France (Beck et al., 2010).

3. The 1992 survey uses four-year age categories (versus five-year starting in 1995).

4. Age categories are larger for the first three surveys (1977, 1981, and 1986)

For prevalence measurements to be consistent across years despite these problems, we need to make assumptions (presented as scenarios) on the variation of prevalence with age within each category for particular age categories and years, which may increase the risk of spurious measures of prevalence.

We followed the steps detailed below to overcome these problems:

- We restricted the age range to 15 to 75 (we drop the information pertaining to the 12 to 15 year-old from the 2000 and 2005 surveys, as well as that pertaining to the 76 to 85 year-old from the 2010 survey).

- We reconstituted prevalence rates for the age categories as defined in the most recent surveys, meaning 12 groups of five years (15 to 19, 20 to 24, until 70 to 74).

- For years for which data are not available along these categories, we used an imputation method based on various forms of linear interpolation and auxiliary data. Since the assumptions made vary by years of survey we present what we did separately for each survey:

  - For the 1995 survey and the 15 to 17 year-old: accurate prevalence rates for youth by sex and by continuous age were published for 1997 based on an auxiliary data source (Baromètre Jeunes 1997-98). These figures show that prevalence for the whole age category 15 to 19 is 81% of that at specific age 18 to 19 for boys and 87% for girls. We therefore use these two ratios to infer the rates for the 15 to 19 from the observed rates for the 18 to 19 in the 1995 survey, separately for boys and girls: $S_{boys15-19,1995} = 0.81 \times S_{boys18-19,1995}$ and $S_{girls15-19,1995} = 0.87 \times S_{girls18-19,1995}$
– 1992 survey: there is no information on those younger than 18 (the first category is 18 to 21). We need to do two things: a) reconstitute prevalence rates for the age categories used in surveys starting in 1995: Prevalence ratios from two age categories are weighted by a set of $\frac{2}{3}$ and $\frac{1}{3}$ coefficients so that the contribution of an observed group to a reconstituted group is proportional to the number of years the observed group contributes to the reconstituted one. For instance, prevalence is observed to be (in 1992) 59.4% for the 27 to 31 year-old men and 48.4% for the 32 to 36 year-old men, and we reconstitute prevalence for men aged 30 to 34 as follows: $P_{r,30-34} = \frac{2}{3} * P_{o,27-31} + \frac{1}{3} * P_{o,32-36}$ = 53.7%, where "o" stands for observed, and "p" for predicted (P are prevalence rates); and b) reconstitute prevalence for the under 18: The approximation detailed above applies to all age categories except the youngest (15 to 19) because we do not have information on the 15 to 17 year-old. The following (rather arbitrary) rule of thumb is applied: $P_{r,15-19,1992} = \frac{P_{o,15-19,1992}}{P_{o,15-19,1992} + \frac{P_{o,20-24,1995}}{Mean[P_{o,15-19,1992}, P_{o,20-24,1995}]}$

– 1977, 1981, and 1986: we needed to reconstitute information on given that age categories are cut below age 18 (for surveys in 1977 and 1981) and age categories are much wider than those of later surveys (for the three surveys). This required the most questionable interpolation assumptions of our imputation work. An improvement on our crude extrapolation method would require using an alternative data source, a large general population survey run by Insee in 1980 (Enquête décennale santé) for access to which we have initiated the request.

Details of the method and calculations used are given in Appendix 1.

2.3.3 Measurement of smoking status

As already discussed, we can find a description of smoking that is stable across all surveys run by INPES. The only requirement is to aggregate regular smokers (one cigarette or more per day every day) and occasional smokers (who may not smoke at all some days and smoke other days) and count them as smokers (non-smokers being therefore respondents who never ever smoke).

This goes against standard epidemiological practice as most studies in that field, and especially studies interested in tobacco-related mortality, interest themselves in regular smokers only. However, since regular smokers make up between 80% and 90% of the total population of those self-reporting as smokers7, this approximation is not too problematic. In our prevalence data by age within

---

7These proportions are calculated for the last four Baromètre Santé, conducted between 1995 and 2010, for which the decomposition is available, see table 3 below, breaking down the population of smokers by regular and occasional, by year, sex and age. Table 3 shows that regular smokers almost always represent more than 85% of all smokers, and close to 90%, for male and females smokers, aged between 30 and 65. Occasional smoking behaviour seems to be more frequent among very young and very old smokers.
cohorts, we only know whether respondents smoke (occasional or regular) or not and some might move from regular to occasional (or the other direction), but we are not aware of any evidence documenting such a trend across cohorts. Overall, grouping together occasional and regular smokers might introduce some noise in our use of prevalence data to simulate MAT, especially if the ratio of occasional to regular varies with time, age, or across cohorts. We use sensitivity analyses to put boundaries around the noise introduced by the definition of smoking as being regular or occasional.

Table 3 here

3 Findings on tobacco prevalence by cohort in France (1975-2010)

3.1 Reconstitution of "pseudo" birth cohorts

Based on the pooling method described above, we can generate age profiles of prevalence for successive cohorts (the prevalence at each age observed for a given birth-cohort across years). We define 5-years birth cohorts starting with the cohort born in 1940-44 and moving ahead by steps of five years (1940-44, 1945-1949, etc.) Since we only have 8 cross sections (spanning 35 years of observation) but 12 age categories (spanning 60 years of age), these age profiles are incomplete and there is not a single generation for which we can describe a comprehensive age profile of smoking prevalence along the life course. Specifically, we can describe prevalence for seven age categories (35 years of age) for five cohorts only (born between 1936 and 1960).

3.2 The average age-profile

Pooling all cohorts together, we can confirm a well-established fact of smoking in France (and most other countries, see, for the case of France, Beck et al., 2010): prevalence starts at one third among teen-agers, then picks up quickly to between 50% (females) and 60% (males) at age 20-24, and starts declining after age 30 in a way roughly linear with age. Female smokers tend to quit earlier.

See graph 1.

Graph 1 here

3.3 Age profiles of prevalence by cohort

Table 4, as well as graphs 2 and 3, present the same age profiles but broken down by cohort.

Table 4, Graph 2 and Graph 3 here

Results for males:
• The assumption that smoking prevalence monotonically decreases after the peak is supported for men as increases in prevalence with age are rare: they are observed for cohort 1926-30 between ages 45-49 and 50-54 (from 44% to 50%), cohort 1936-40 between ages 45-49 and 50-54 (from 40% to 45%), and cohort 1946-50 between ages 40-44 and 45-49 (from 41% to 47%).

• All profiles look similar, which suggests a common age profile of initiation and cessation for all cohorts, without much change in terms of when more recent cohorts tend to quit smoking.

• The main difference across generations is that more recent generations start at a lower level of prevalence: The lowest prevalence rates at age 15-19 are observed for cohorts 1986-90 and 1976-80, with prevalence just above 30%. Conversely, almost 45% of cohort 1961-65 and still 40% of cohort 1956-60 were smoking at this early age.

• Similarly the peak at 20-24 is reached at lower levels for more recent cohorts: more than 60% of males born before 1970 were smoking when they were between 20 and 29 years-old, even 65% for cohort 1961-65 and up to 70% for 1956-60. But it is down to 50% for cohorts 1986-90 and 1976-80.

It is therefore clear that, for males in France, the decline in prevalence of smoking in recent years is due to fewer individuals (youth) starting smoking rather than more smokers quitting earlier.

Results for females: they are more complex than for males, due to large temporal shocks in the 1970s and 1980s affecting all age categories and cohorts.

• Profiles vary more erratically for women than for men, differing markedly from the average tendency shown on graph 1. In particular, we can observe upturns in tobacco consumption when women get older in several cohorts. Since such a phenomenon cannot be observe in the same proportions for men, where these cases are very uncommon, and it cannot be an artifact due to data. Is that increase due to relapse smokers (eg smokers who have quit for some years and then returned to smoking) or to late entry of new smokers? We cannot disentangle these two interpretations based on the data at hand. This finding could partly undermine our key assumption that smoking abstinence once attained will be sustained ever after. But the alternative interpretation is that women from older cohorts, who were prevented from starting smoking before age 20 when social norms strongly opposed smoking among women, were able to do so later (in the 1970s and 1980s, when they were older), at a time when female smoking was better tolerated. The latter interpretation would be innocuous to our modeling of the effect of prevalence on mortality.

• For women born between 1936 and 1945, the increase in prevalence takes place at older ages, corresponding to the period between 1990 and 1995.
For that cohort, prevalence increases from 10% only at ages 50-54 to almost 18% at ages 55-59. The same can be observed for women born in 1941-45, with prevalence increasing from 14% at 45-49 to 19% at 50-54.

- Similar increases can be observed at younger ages for cohorts 1946-50 and 1961-65 in particular. For those generations the upturn takes place around ages 30 to 35.

- Prevalence has increased, on average, for women born during the 1950s and 1960s compared to previous generations, but seems to be on the decline for most recent ones: the proportion of smokers is around 45% at ages 30-34, around 40% at ages 35-39, and still between 35% and 40% at ages 40-44 for cohorts 1951-55 to 1966-70.

- For women born after 1970, prevalence decreases to 40% at ages 25-29 for cohort 1971-75.

- Finally, tendencies for the most recent two generations (women born between 1981 and 1990) are inconsistent and hard to make sense of. Women born between 1981 and 1985 show very high prevalence rates with no sign of decline with age (prevalence is at 45% from age 15 to age 29) and, for the 1986-1990 cohort, prevalence starts low (30% at ages 15-19) but increases dramatically to 45% to reach the same level of consumption as the previous cohort. This might suggest the reappearance of an upward trend in smoking behaviors for current and next generations of women but the jury is still out for lack of evidence on the prevalence for these cohorts later in life.

3.4 Measuring changes in the age pattern of quitting behaviour across cohorts

We now turn to the question motivating the study of mortality attributable to tobacco: even though eye ball comparisons presented in the previous section clearly show that most of the reduction in overall prevalence observed in the 2000s in France is the result of lower rates of youth initiation and lower peak prevalences around age 25 (at 20-24 or 25-29, depending on cohorts\(^8\), we can still detect some action around the slope of decreasing prevalence with age after 30. This suggests that, although dwarfed by the much larger effect on the initiation and peak levels, an increase in the quitting hazard ratio may also have taken place in response to the vigorous push against smoking in the late 1990s and early 2000s (strong tax increases and smoking bans in public place implemented by the Jospin and Raffarin governments).

\(^8\)Two cohorts do not follow this general pattern: men born in 1976-80 (peak at 30-34), and women born in 1946-50 (peak at 30-34 on the latter, see section 5.1; we suspect some statistical artifact and will use smoothed data for this cohort thereafter). For these two cohorts, we still standardize to their rate at 25-29, which explains why we see two rates greater than 1 on these graphs.
To do this, we standardize cohort age profiles, attributing a value of 1 to the peak prevalence and comparing prevalence by age to that peak value for each cohort. This way, changes in the absolute level, reflecting changes in the probability to start smoking, are neutralized, and we can test how much has changed in cessation patterns. As already noted, seven cohorts only are observed over a long enough time for such an exercise to be meaningful. Results are presented in graphs 4 and 5, where the peak can be easily discerned as the value 1 for each generation.

3.5 Findings on standardized ratios for male smokers

Of two male smokers at 20-29, at least one will have quit before age 50. The quit hazard ratio for male smokers does seem to have increased slightly at age 45-49 across cohorts: At age 40-44, we can observe five cohorts (1946-50 to 1966-70) and the quitting rate increases from 30% for cohort 1946-50 (almost one quitter at that age for three smokers at peak age) to close to 40% for the next three cohorts (1951 to 1960) before dropping again at 25% only for the most recent cohort (1961-66). At age 45-49, the quitting rate increases from cohort 1946-50 (35%) to cohorts 1951-60 (45%) before dropping to 40% (still higher than the oldest cohort) for cohort 1961-66. However, for more recent cohorts, we cannot observe any movement toward higher quitting hazard ratios before age 40.

3.6 Findings on standardized ratios for female smokers

40% at least of female smokers at 20-29 will have quit before age 50 (50% for the two oldest cohorts and 40% for cohort 1956-60). The picture is less clear than for males for cessation after 40: At age 40-44, we can observe five cohorts (1946-50 to 1966-70) and the quitting rate first increases substantially from 20% for cohort 1946-50 to 40% for the next three cohorts (1951 to 1960) before dropping at 25% only for the most recent cohort (1961-66). At age 45-49, we lose that most recent cohort (and are left with four as a result): the quitting rate increases dramatically from cohort 1946-50 (25%) to cohorts 1951-60 (45%) before dropping to 40% for cohort 1961-66. At this age (40-45) at least there seems to be a trend toward more quits among females, a sign that maybe anti-smoking policies or information on smoking may have convinced more smokers to quit before it is too late. At age 50-54, three cohorts can be compared (1946-60), and the rate is stable at 50%. Policies and information would therefore help some smokers who would have quit at 50 to quit sooner in life, which may have an impact on their life expectancy as discussed in the next section. For the 1971-75 generation, the pattern slightly differs from the rest with a quick and major withdrawal from smoking around age 30. It is confirmed that the youngest generation of women (born in 1976-80) have not only a high level of consumption but also, apparently, a lower propensity to quit. We substantiate here our previous finding of a large discrepancy in smoking behaviors between
the most recent two generations (1976-80 vs. 1971-75) in terms of absolute level and relative decline along life. This trend would, of course, need to be confirmed in the future.

4 A new method to reassess MAT

We now describe how we propose to calculate MAT in a way that will allow us to separate the effect of duration in smoking and duration from cessation from the effect of total prevalence. This is very different from what is currently done for calculating MAT (in France or in any other country).

4.1 Epidemiological evidence on MAT

Within the vast literature on the effect of smoking on morbidity and mortality, much more is to be found on the underlying biological model linking tobacco consumption to the risk of cancer (Jha, 2009) than on the epidemiological measure of the effect of duration, daily dose, and cessation of smoking on mortality.

Section 1.1 discussed the few studies on duration(s) and MAT and their methods and findings on relative risks or absolute mortality as a function of duration (number of years smoking), intensity (number of cigarettes smoked per day) and, in some cases, years since cessation for ex-smokers. No epidemiological study calculates relative risks for both current smokers (according to duration) and ex-smokers (according to duration from cessation). Studies show that duration (in smoking or since cessation) is the most important determinant of mortality, much more than intensity. Age at initiation has no effect on mortality, independent of duration.

4.1.1 Overview of recent epidemiological studies

We summarize here the salient characteristics of these studies, their strengths and limitations, in order to select the estimators of risk as a function of duration and other factors that we can use in our modeling. The overview is provided in table 5. For completeness, we mention the seminal paper of Doll and Peto (1978) even though we do not use their estimator (it is based on a very specific population of British medical doctors).

All these studies differ in scope, methodology and robustness of findings. No study can provide a comprehensive and perfectly accurate framework enabling to calculate mortality caused by smoking for all causes of death and that would be applicable in all contexts.

Table 5 here

4.1.2 What can we get from these studies?

In many respects, Flanders et al. (2003) appear to be the best choice for our purpose:
In fighting smoking, would later not be better than sooner?

- The scope of the study is compatible with our expectations, since it is focused on the US (clinical/epidemiological background comparable to France) for both men and women aged from 40 to 79 (the age people are at higher risk of lung cancer).

- It is robust in terms of data used and statistical methodology: CPS-II is one of the reference surveys in the field and can be considered recent enough, especially compared to other studies. It has a large sample size of more than 100,000.

- The empirical model is straightforward even though it includes crude additional adjustment of potential co-founders. The empirical model allows us to disentangle the effects of duration and intensity (but not of duration since cessation).

- Finally, results are conveniently expressed: influences are expressed as coefficients of a model of the absolute probability of dying from lung cancer instead of risks (relative to the baseline risks of never smokers). Using relative risks in a simulation of MAT is not without problems, most notably the fact that these risks are relative to a baseline risk (without the exposure to smoking) which is not easy to estimate in studies, even large studies like the CPSII, because it is a very small one (close to 0). As a consequence, the smallest error in the baseline can lead to very large differences in the estimate of the RR, and, as a result, in the total number of deaths attributable to smoking.

The main results of Flanders et al. (2003) can be described as the following eight different equations of mortality referred to in section 1.1 of this report (respective influence of duration and intensity are estimated (under exponential form) separately for 4 age groups and for men/women):

- Men, 40 to 49 years old: \( \hat{M}_m = e^{-17.9}D^{1.9}I^{0.95} \)
- Women, 40 to 49 years old: \( \hat{M}_w = e^{-20.2}D^{2.8}I^{0.96} \)
- Men, 50 to 59 years old: \( \hat{M}_m = e^{-17.4}D^{2.6}I^{0.52} \)
- Women, 50 to 59 years old: \( \hat{M}_w = e^{-17.2}D^{2.2}I^{0.75} \)
- Men, 60 to 69 years old: \( \hat{M}_m = e^{-15.7}D^{2.4}I^{0.37} \)
- Women, 60 to 69 years old: \( \hat{M}_w = e^{-14.1}D^{1.5}I^{0.78} \)
- Men, 70 to 79 years old: \( \hat{M}_m = e^{-13.0}D^{1.8}I^{0.39} \)
- Women, 70 to 79 years old: \( \hat{M}_w = e^{-13.2}D^{1.3}I^{0.95} \)

Other studies can also be helpful to some extent, but compared to Flanders et al. (2003) they generally appear too restrictive, limited in their methodology or simply too old. For instance, Streppel et al. (2007) is only about men and
based on a sample size of 1,400. Knoke et al. (2008) is based on CPS-I which took place 50 years ago. Finally, Rachet et al. (2004) although proposing a very robust methodology is based on a small sample and provides RRs only as its main results.

Like others, the work of Flanders et al. (2003) is not without limitations:

- Only lung cancer is considered as a morbidity consequence of tobacco consumption. This choice is very common, as it can be seen in the scope of studies selected here, but it is insufficient when it is about measuring the overall deaths caused by tobacco.

- No particular attention is paid to ex-smokers and to the impact of past smoking on the incidence of lung cancer.

These two potential limitations are examined in the next subsection.

4.2 Details of the method to estimate MAT in France

We present here the steps involved in calculating MAT. First, however, we summarize and discuss the assumptions we make. Some of them have already been presented, other are specific to this calculation. In particular, a great deal of attention is paid to the inclusion of ex-smokers in the calculation.

4.2.1 Lung cancer as a proxy for overall MAT

Using Flanders et al. (2003) and running calculations on that basis implicitly relies on the following equivalence: $MAT \simeq$ mortality caused by lung cancer, which is not perfectly accurate, for the following reason: Lung cancer is not the only cause of death due to smoking. It is certainly the most direct and unquestionable one and it is mostly caused by tobacco (the risk for non smokers is very low), which is why it is commonly used as a proxy in most studies. Peto et al. (1992) recommend calculating the risk of dying for diseases other than lung cancer based on calculations of specific population attributable fractions for every disease possibly caused by cancer (see section 1.1 where the method is described in detail). An alternate rough approximation would be to consider that overall MAT is a linear function of specific lung cancer MAT ($MAT \simeq \alpha \times MAT_{lung}$). We could then extrapolate our results of the calculation of lung cancer mortality to overall MAT if we can find a plausible estimator of this factor $\alpha$ in the literature.

4.2.2 Impact of ex-smoking habits on mortality

Flanders et al. (2003) do not estimate the mortality of ex-smokers. Yet, it is obvious that ex-smokers have higher risks of dying of lung cancer than never smokers. Considering otherwise would mean that the excess risk of cancer instantly reverts to 0 when someone quits smoking.

Knoke et al. (2008) conclude that the magnitude of this excess risk depends both on time since cessation and age at cessation, but that the number of
cigarettes smoked while smoking does not matter as much. The risk of lung cancer is reduced by half after 7 years if the smoker quits at 40 and after 9 years if he quits at 50. The risk is even reduced by 90% after 15 years when a smoker quits between 30 and 40. Quantitatively Knöke et al. (2008) can be summarized in an equation linking the decrease in excess risk (denoted \( f \)) to age at cessation (\( A_c \)) and time since cessation (\( T_c \)): \( f = e^{-0.274 - 0.00279 A_c + (T_c - 2)} \)

This equation will be used in our method as a complement to account for lung cancer mortality of ex-smokers. The method to achieve that is as follows:

1. We consider that 30 is the pivotal age between smoking initiation and cessation. Our working assumptions are that no one starts smoking after 30 and that no one quits before 30.

2. To assess the magnitude of smoking cessation, we use long-term prevalence data as they have been reconstructed in our baseline methodology (scenario 1). If \( X\% \) of a birth cohort smokes in year \( N \) and \( Y\% \) smokes at \( N + 5 \), we just consider that \( (X - Y)\% \) quit smoking in the meantime. We then neglect possible in/out flow considering that no one resumes smoking when he has quit earlier in life (and that no one starts after 30).

3. We are interested in persons at risk of lung cancer only, aged 40-79 in 2010. Then only individuals born between 1931 and 1970 are entered in our calculation.

4. Arbitrary interpolation rules are necessary to transform our set of discrete measures (survey years and age categories) into continuous values for the sake of applying the equation derived from Knöke et al. (2008): We use the middle age of intervals as the average age of cessation. For example, for the fraction of a generation who is smoking at 40-44 and has quit at 45-49, all people are said to have quit exactly at 47. For time since cessation, the situation is similar. Imagine people who were smoking in 2005 but no longer in 2010: it seems natural to suppose that they have quit for 2.5 years on average. The same idea applies for more long-term abstinence which leads to a linear form of \( 2.5 \times 5x \) for time since cessation. People who quit between 2000 and 2005 are all considered as having quit since 7.5, people who quit between 1995 and 2000 are all considered as having quit since 12.5 years, etc.

5. Of course, age at cessation and time since cessation are not independent parameters for a same birth cohort. Consider for example the 1966-70 cohort (aged 40-44 in 2010). Among those, when we are interested in identifying ex-smokers, three situations only are possible: individuals who quit at 30-34, at 35-39 or lately, between 40 and 44. Considering how our data are organized, time since cessation is automatically deduced from that: people have quit respectively for 12.5 years, 7.5 years or 2.5 years in that case. Overall, we consider 45 different combinations of values of interest for the parameters (age at cessation and duration since cessation),
which makes a sort of typology of ex-smokers. They are represented in table 6 below.

6. We need to estimate the percentage of persons who quit smoking for each cohort in our range (1931-70) and at every moment of their life (from 30 to 79). Note that we cannot evaluate all cessation rates, in particular cessation at older age 75-79 for the first cohort (1931-35), due to lack of data. To estimate these percentages of cessation, we first need to identify the peak of consumption (highest prevalence rate) for each birth cohort. The peak always takes place between 20 and 29 (at least for the cohorts we are interested in), so it seems straightforward to take the maximum value between prevalence at 20-24 and prevalence at 25-29 as the reference peak of prevalence. Unfortunately, prevalences at youngest age categories are no longer available for cohorts born before 1946. In this case, the reference value is chosen as the prevalence for the youngest category available for a given cohort (prevalence at 30-34 for cohort 1941-45, at 35-39 for cohort 1936-40 and at 40-44 for cohort 1931-35). The cessation rates can then be calculated by subtracting the prevalence rate at a given age from the prevalence rate of the previous age category. There is a peculiarity for the age category 30-34 which is compared to the peak of prevalence (which can be either for 25-29 or 20-24). For a given birth cohort $k$, and a given age group $j$, if $P$ denotes the prevalence rate, $Q$ the percentage quit (the proportion of cohort $k$ who quits smoking at age $j$) can be calculated as: $Q_{k,j} = \text{MAX}[P_{k,20-24}; P_{k,25-29}] - P_{k,j}$ for $j = 30-34$, and $Q_{k,j} = P_{k,j-1} - P_{k,j}$ otherwise.

7. There is one last problem though. Contrary to our assumptions and expectations, it happens that smoking prevalence rises when a cohort gets older. We set arbitrarily quit rates at 0 in these cases, considering that no one quits smoking in this cohort between at these ages.

**Table 6 here**

### 4.2.3 Steps of the calculation

All our methodological choices and key assumptions directly shape the results. We recapitulate all of these assumptions in the table below, setting default values for the potentially varying parameters.

**Table 7 here**

To summarize:

- we use 15 as the age at initiation (an assumption that is supported by empirical data for France, e.g., Beck et al., 2010); it is to be noted that Flanders et al. (2003) use a surprisingly high value of 22.5.

---

9Among the 45 different combinations, such situation happens 7 times for men and 10 times for women, as already described. These situations correspond to the unexpected increases in curves of graphs 4 and 5.
IN FIGHTING SMOKING, WOULD LATER NOT BE BETTER THAN SOONER?

- we use a constant value for intensity of smoking (20 cigarettes a day for everyone); alternatively we could let this value vary by sex and age group using recent survey data (ESPS 2010 for instance).

- the crucial parameter â€œduration of smokingâ€ is naturally derived from the series of prevalence rates by age recreated above and assumption 1 that all start at age 15. Alternatives can be imagined using either substantial enhancements in data extrapolation or fictional rates of smoking for current and future generations. Hypotheses about cessation have been thoroughly discussed above.

The steps are as follows:

1. Calculate the individual mortality risk of lung cancer based on the equations in Flanders et al (2003). It is worth noting that their equations have to be duplicated for some age groups, since they are provided by 10 year age categories, while we use 5 year age categories for our prevalence estimates. Our choice is to apply the same equation (say for men aged 40-49) to both categories 40-44 and 45-49.

2. Calculate the number of current smokers at risk in 2010. We extrapolate the prevalence rates (regular + occasional) at the latest known date (2010) separately for men and women aged 40 to 79 to the size of the French population in 2010 (as provided by the census for 2010 in suitable 5-years age categories by Insee). Note that here, and unlike in the reconstruction of long-term cohorts, we use the prevalence rate for the 75-79 as published in the 2010 survey to populate this category.

3. Calculate the number of deaths due to lung cancer for current smokers. This step directly results from steps 1 and 2: we apply the individual risks of dying of lung cancer to the population of current smokers, separately by sex and age.

4. Calculate a risk for ex-smokers, as a function of time since cessation, based on the results in Knoke et al. (2008). The combinations of possible values for the two parameters â€œage at cessationâ€ and â€œtime since cessationâ€ lead to 45 different values of decrease in the risk of lung cancer mortality compared to the baseline risk of the individual at the time he quit (recall that, since all start at age 15, we know duration in smoking from age; here, we infer completed duration as smoker from age at the time of quit). Applying these values to the the probability that the smoker would die of lung cancer (incidence, equivalent to mortality in the same year) at the time they quit we can calculate probabilities of lung cancer of ex-smokers according to duration since quit.

5. Calculate the number of ex-smokers who stopped by age at cessation and time since cessation. It is necessary to start with applying the method described in detail in subsection 4.2.2 to estimate the percentage of cessation
for each birth cohort and at every age. To extrapolate to numbers in the population, we apply these percentages to the living population gauged in censuses for the different points of measure (long-term series of censuses from 1975 to 2010 provided by Insee). For example, men aged 40-44 were 2.2 million in France in 2005 (according to the census); we estimate that around 10% of individuals in the 1961-65 cohort quit smoking in 2005 (our calculations based on prevalence data), therefore 220,000 men born in 1961-65 are said to have quit tobacco at 40-44. In other words, in 2010, 220,000 individuals have quit smoking with age at cessation 42.5 and time since cessation 7.5 years. We implicitly assume that all of these 220,000 individuals are still alive in 2010 and potentially at risk of lung cancer.

6. Calculate the number of deaths due to lung cancer for ex-smokers. In the last step, we sum up the number of ex-smokers calculated in step 5, aggregating these numbers by sex and age category (and still assuming they all made it until 2010). Applying the specific risk of mortality retrieved from step 4 allows us to calculate the number of deaths among ex-smokers in 2010.

5 Results

This section:

1. Presents our estimation of MAT for the year 2010, based on current observed prevalences by cohort and the parameters derived from the epidemiological literature for mortality as a function of intensity, duration, and time since cessation.

2. Discusses the sensitivity of results to the values of the main parameters.

3. Generates a central tendency of future MAT, under the assumption that smoking behaviours remain fixed in future cohorts (at the level and age pattern observed for current cohorts).

4. Presents the results of "What if" scenarios in which some key parameters of the age profiles of prevalences of smoking are altered separately for different cohorts to a) generate realistic changes in overall prevalence (at the population level) and b) reflect extreme situations in which changes take place mostly through initiation (starting smoking is less common) or cessation (quitting is more frequent).

5. Summarizes key assumptions and discusses potential methodological improvements.
5.1 Mortality Attributable to Tobacco in France under reference assumptions: a reconstitution

As a reminder, our main assumption are as follows:

1. No one starts smoking after 30.

2. No one quits before 30.

3. Once smokers quit (for more than five years, our observation period) they never relapse. We do not try to distinguish occasional from regular smokers and will apply the same mortality hazards to both sub-populations of smokers.

4. Since epidemiological studies measure lung cancer mortality only, we predict mortality for that cause on the basis of prevalence rates by age and cohort.

The general principle of our reconstitution of MAT is straightforward: MAT is a function of intensity and duration of smoking for current smokers (based on Flanders et al., 2003) and of time since cessation for ex-smokers (based on the relative decrease in risk compared to risk at cessation calculated in Knode et al., 2008).

We assume that intensity is the same for all smokers but we run various simulations using different values for this parameter. The central scenario uses a value of 20 cigarettes per day but sensitivity analyses allow for other values. We also assume that age at initiation remains fixed for all cohorts, at 15, but, again, run sensitivity analyses around this value.

Under these central values for intensity and age at initiation, our simulation produces 23,636 deaths caused by lung cancer for the year 2010, comprised of 15,635 deaths for males and 8,001 for females, and 11,015 current smokers dying versus 12,621 ex-smokers. According to the Centre d’épidémiologie des causes de décès (CépiDe), affiliated with the INSERM (French National Institute for Health and Medical Research), there were 30,882 deaths caused by lung, larynx, trachea and bronchi in 2009. It is not easy to compare our simulated number of deaths due to lung cancer only to an observed number aggregating other causes of death to lung cancer (we would need to know a good value for the share of lung cancer among those four in France) but, most likely, our method slightly under-estimates the true number of deaths for lung cancer. This is not surprising since not all lung cancers are caused by smoking and we reconstitute deaths caused by lung cancer as a result of smoking only. Detailed results of our reconstitution by sex, age category and smoking status in 2010 can be found in the following.

Table 8 here
5.2 Sensitivity analysis: how is our estimate of MAT influenced by the assumptions made to impute missing data as well as the values chosen for our key parameters?

Data imputation (reconstitution of age categories not found in publications for years prior to 1985): We run alternative scenarios on smoothing/extrapolation and find a range of values for total MAT of [23,377 - 24,536] representing a variation of -1.1% to +3.8% compared to our reference scenario. This suggests that our estimate is not sensitive to the assumptions we made to impute missing data.

Results are more sensitive to the value used for duration and intensity: using intensity as reported in the EOPS survey for 2010 (and allowing it to vary with age and sex as reported in the survey) leads to 19,200 deaths caused by lung cancer only in 2010 (13,630 males and 5,570 females, with 8,980 deaths of current smokers compared to 10,220 deaths of ex-smokers). This represents a decrease of -19% compared to the reference scenario.

Graph 6 summarizes the effects of these two parameters (intensity and age at initiation): for a given age of initiation, MAT increases almost linearly with intensity. For example, if all smokers start at 18, an intensity of 10 cigarettes a day would yield approximately 14,000 deaths caused by lung cancer. Starting at 15 with intensity of 15 cigarettes per day or initiation at 18 with intensity of 20 cigarettes per day yield the same result of about 20,000 deaths due to lung cancer.

Graph 6 here

5.3 How will MAT evolve from 2010 to 2060?

We now use our simulation method to generate MAT until 2060, under a variety of hypotheses regarding prevalence: we start with a projection based on smoking prevalences by age for future cohorts reflecting the trends in prevalence observed for previous cohorts. We then construct "What if" scenarios in which a shock takes place on prevalence, either on initiation or on cessation.

Most recent trend in prevalence assumed to hold in the future

We need to generate smoking prevalence for all cohorts and all age categories between 15 and 79 for the years 2010 to 2060. We keep the time trend observed in recent years through generating these prevalence rates as the result of recursive smoothing of the rates for the three previous cohorts:

1. Prevalence at initiation age for cohort $Y$: $I_Y = \frac{I_Y - 5 + I_Y - 10 + I_Y - 15}{3}$

2. We generate prevalence rates for age $[x; x+5]$ and cohort $Y$ as the product of the prevalence for the same individuals five years before (when they were aged $[x-5; x]$) by a rate of increase between these two age categories for that cohort. For instance, prevalence at age 20-24 is based on the initiation
rate for the same cohort five years prior, and then generate prevalence at age 25-29 five years later, applying a rate of change for that cohort:

\[ P_{Y,[20:24]} = I_Y \times (1 + C_{Y,[20:24]}) \quad \text{and} \quad P_{Y,[25:29]} = P_{Y,[20:24]} \times (1 + C_{Y,[25:29]}) \]

and so on and so forth.

3. The rate of change for a given age category and a given cohort \( C_{Y,[x:x+5]} \) is based on the average of observed/predicted rates of change at the same age for the 3 previous cohorts:

\[ C_{Y,[x:x+5]} = \frac{C_{Y-5,[x:x+5]} + C_{Y-10,[x:x+5]} + C_{Y-15,[x:x+5]}}{3} \]

Table 9 below provides prevalence rates for some cohorts and age categories, generated by the method described above.

**Table 9 here**

Once we know the prevalence rates by age, we use population projections by age provided by INSEE to generate distributions of smokers and ex-smokers in the future, by age, duration of smoking and durations since cessation. We then apply our the effects of duration as calculated by Flandens et al., as we did for 2010 (see above). Graph 7 below provides our estimates for MAT from 2010 to 2050, assuming our central values for the main parameters (initiation at age 15 and daily dose of 20 cigarettes per day).

**Graph 7 here**

Under the reference scenario, the number of deaths per year caused by smoking will increase over the next 30 years, up to a maximum value of 31,600 deaths around 2035. This increase from 2010 to 2035 rate (+6% every 5 years on average) mainly reflects the fact that the generations born during the 1930s and 1960s, who were heavy smokers, will be at maximum risk of lung cancer within the next 20 years. After that, MAT will slightly decrease and stabilize around 30,000 deaths. This projection is certainly largely artificial though, due to the method we use to generate future prevalences for future cohorts (since predicted prevalences are recursively based on average values of the 3 previous cohorts, at some point prevalences tend to stabilize across cohorts).

**Under alternative scenarios: shocks on prevalence**

We now introduce shocks in which prevalence changes abruptly, either due to an increase in cessation rates (seasoned smokers suddenly quit en masse) or a marked decrease in initiation among teenagers. In other words, we answer the following question: how would mortality due to tobacco change if public health policies and/or quitting programs proved to be effective? To assess the impact of these hypothetical changes, we test 2 kinds of scenarios for the future and simulate MAT under each scenario:

1. Smoking initiation decreases drastically by a factor 0.1, from 2015 on for each new cohort. Let us follow the next birth cohort, exposed to initiation in 2015 (born in 1996-2000, who will be aged 15-19 in 2015). If nothing happens, their tobacco initiation rate can be estimated at 34.6% for males.
and 35.3% for females (average of initiation rates for cohorts 1991-95, 1986-90 and 1981-85). In the present scenario, we assume instead that real initiation is lowered by a given coefficient \( L \): 

\[
I_{\text{Males}} = 0.346 \times (1 - L) \\
I_{\text{Females}} = 0.353 \times (1 - L)
\]

The same coefficient applies to all cohorts born after 2000, and their initiation rates under this scenario are therefore those calculated under the reference scenario above reduced by this factor \( L \). The shock is a one-time shock, not a cumulative one. Prevalence rates at older ages are calculated as in the reference scenario, applying unchanged cessation rates by cohort and age to this lower initiation rate of each cohort.

2. Cessation rates increase by \( Q \) percent at all ages. Again, this is a one-time shock, not a cumulative increase in cessation rates. This can be thought of as the result of a new treatment helping \( Q \) percent more smokers to quit at each age.

We chose these two scenarios for their simplicity: they can be expressed as the effect of an effective public health campaign reducing initiation rates among future cohorts by a factor \( L \) for ever or of an effective treatment increasing quit rates by \( Q \) percent for ever. Table 10 below provides examples of resulting smoking prevalence rates for two birth cohorts (1991-95 and 1995-2000) under three scenarios: the reference (central) scenario in which temporal trends apply to future cohorts; a quite successful public health campaign, with a decrease in initiation by 50%; and a successful treatment increasing quit rates by 50%. Of course, more complex scenarios (e.g., in which the initiation rate would decrease more progressively across cohorts) would be more realistic, but it would add complexity without affecting our main message. These scenarios are "what if", not projections, they are meant to simulate the effect of a shock in one of the determinants of smoking prevalence only (initiation or cessation), and, as a result, will tell us where we should invest our next Euros (prevention or treatment) if our objective is to reduce mortality attributable to tobacco.

**Results of the scenarios with various values of \( L \) and \( Q \):**

We test our two scenarios with a range of values for \( L \) and \( Q \): \( L \), the initiation coefficient, starts at 0.25 (prevvalence at initiation for future cohorts is 75% only of the latest observed rate), and takes values of 0.50, 0.80, and 1.0 (a highly hypothetical scenario in which, all of a sudden, nobody starts smoking in future cohorts). For \( Q \), the quitting coefficient, we start at 0.25 (quit rates increase by 25% from latest observed cohort to the next) and use values of 0.50, 1.00 (quit rates double) and 2.00 (quit rates are multiplied by 3 at all ages). No scenario for quitting is as extreme as our "no initiation" scenario, as that would imply full cessation at a given age. This would not only require the discovery of a new treatment, but also some kind of drastic regulatory change motivating all smokers past a given age to take the treatment. We are content here to simulate a new therapy helping smokers who want to quit to succeed, assuming the regulatory environment remains constant. It is important to keep in mind, though, that our extreme initiation scenario is absolute whereas our
extreme cessation scenario remains relative (some smokers never quit). Results are expressed as number of deaths due to lung cancer each year between 2010 and 2060. Graph 8 shows MAT per year in case of a public health shock, with varying values of L: the effect of reduced initiation on mortality is nil before 2035 and the total number of lives saved (MAT prevented) over the 50 year period of our simulation varies between 8,000 (with a 25% reduction in initiation) and 32,000 (in the extreme scenario of no initiation at all in future cohorts). Graph 9 shows the same results but in the new treatment scenarios: the effect on MAT starts in 2010 and total lives saved in this scenario vary between 20,000 (quit rates increase by 25%) and 75,000 (quit rates triple) over the course of the next 50 years.

**Graph 8 and Graph 9 here**

These crude scenarios clearly show that increasing cessation rates seem to be an effective strategy for the coming 50 years: if we find a treatment that doubles quit rates at all ages, we will prevent between 4,000 and 7,000 deaths (out of 24,000 to 30,000 attributable to tobacco) each year. If we prevent the youth from starting smoking at all, we will prevent 16,000 deaths in 2060 but almost none before 2040. Using a time discount rate of 5% (a life saved next year is worth 5% less than a life saved today), these figures indicate that the present value of deaths prevented by a doubling in the cessation rate would be ten times the present value of deaths prevented by cutting initiation in half. We could therefore invest much more of our Euros into trying to find a good treatment for addicted smokers to get the same result as a campaign aimed at preventing the youth from starting smoking.

Of course, in the long run (more than 50 years), preventing initiation pays off relative to increasing cessation. One scenario we have not tested so far, but which might be of interest, is a new treatment with a very high rate of success for hard core smokers (smokers over the age of 50, say); rather than multiplying quit rates by a coefficient, this new scenario would impose a success rate of 80% or 90% for all smokers past age 50. Such a scenario would guide us in understanding better reimbursement strategies for the cessation treatment: at what age should we start covering it to maximize the effect of public expenditures in terms of prevented deaths (and assuming coverage matters for success)?

This is a first step toward a better understanding of the determinants of MAT (and, possibly, of better point estimates of MAT). What is left to be understood and described now are the various selection effect at play in smoking behaviours: if marginal smokers are more likely to be prevented from starting smoking by public health campaigns and if these marginal smokers are also more likely to quit before it is too late, then preventing initiation may be even less efficient than we predict here. Similarly, we need to better understand the relationship between quitting and mortality risks: if healthier smokers quit first, increasing quit rates will be more efficient than predicted here. But if healthier smokers do not quit, precisely because they are healthy, increasing quit rates might be useless.
References


Appendix 0: The WHO global report 2012: Mortality Attributable to Tobacco

Direct tobacco smoking is responsible for 5 million deaths a year (and 600,000 from second-hand smoking) and is expected to kill 8 million people a year in 2030 if nothing is done. To prevent this, WHO introduced a tobacco control package, known as MPOWER:

- M: Monitor tobacco use and prevention policies - the WHO has developed a framework of national surveys on prevalence of smoking but is now working on expanding it to monitoring the effect of smoking on mortality and morbidity.
- P: Protect people from smoking
- O: Offer help to quit
- W: Warn about dangers
- E: Enforce bans on advertising, promotion and sponsorship by tobacco firms
- R: Raise taxes on tobacco

In this report, WHO estimates mortality attributable to tobacco (MATT) for the world, WHO regions, and particular regions (when data are available). They use the PAF (population attributable fraction) approach: if we know P, prevalence of a risk factor (smoking) and RR, relative risk of dying for those with compared to those without the risk factor, the surplus mortality due to the risk factor (number of deaths that happened that would not have taken place without the risk factor, i.e., if the risk factor were nil) is P.(RR-1), and baseline risk of dying is P.RR + (1-P) = P.(RR-1)+1 (mortality risk being 1 among those without the risk factor). Therefore, the proportion of total deaths that can be linked back to the risk factor is: 

\[ PAF = \frac{P \cdot (RR-1)}{P \cdot (RR-1) + 1} \]

PAF is 0 when either P=0 (no prevalence of the risk factor) or RR=1 (no risk associated with the factor beyond baseline risk) and tends toward 1 when P=1 AND RR is infinite (very large). Multiplying the number of cause-specific deaths by the PAF for that cause of death (e.g., lung cancer) provides MAT for that cause. Adding over all causes of death yields total MAT.

5.4 Estimating prevalence of smoking, P

Ideally, one would like to have prevalence data by age (duration) in order to take into account the fact that relative risks (RR) change with age (reflecting duration). However, such data do not exist in most countries and the World Health Organization (WHO) relies on the Peto strategy: derive prevalence from lung cancer mortality, and use such derived prevalence to infer MAT for all
causes. The method uses mortality due to lung cancer in the population, $C_{LC}$, or, rather, the excess mortality due to lung cancer in the population compared to baseline mortality from lung cancer among non-smokers: SIR is a proxy for $P$, where

$$SIR = \frac{C_{LC} - N_{LC}}{C_{LC} - N_{LC}}.$$  

Of course, since $N_{LC}$ is not known (otherwise the rate for smokers would be known), SIR is estimated as follows:

$$SIR = \frac{C_{LC} - \hat{N}_{LC}}{C_{LC} - \hat{N}_{LC}}.$$  

### 5.5 Estimating the relative risk of dying, RR

They use estimates derived from the CPS-II study. The RR for lung cancer are taken as is, but RR for other causes of death are divided by two, as RR provided by CPS-II do not control for possible confounding factors. Interestingly, some RR are calculated at various ages and do not only represent the effect of duration of smoking (because RR actually decrease with age: among really old people, smoking is not much of risk relative to an already high baseline; this is the case for ischemic heart disease and stroke, where the RR is about 3 before age 60, but falls under 2 after.)

Appendix 1: Calculations to extrapolate prevalence rates into detailed age categories for survey years 1985, 1980 and 1975

The following approximations are used to reconstruct accurate prevalences by age-category based on results from 1986, 1981 and 1977 surveys. In some cases, observed prevalences in surveys can be directly used as values for estimated prevalences, especially when age categories are very close, differing only by boundaries. But in most cases, we calculate weighted prevalences, using the structure into sub-categories as it can be observed in closest year of collection data. Let’s consider an example to illustrate this strategy:

First, consider year 1986 and the prevalence observed for men aged 25-34 (noted T). We want to refine this prevalence into two different figures for the sub-categories we are interested in, namely 20-29 and 30-34. Notation: $P_1^t$ for the 20-29 and $P_2^t$ for the 29-34. Moreover, we know the prevalence rates for these exact age groups for the year 1992 (the closest year we observe), that we note $P_1$ and $P_2$. $P_1^t$ and $P_2^t$ are solutions of two equations:

1. the average of estimated prevalence rates must be equal to the observed rate for the wider age category observed in the same year: $\frac{P_1^t + P_2^t}{2} = T$
2. By assumption, the ratio of estimated prevalence rates within the age category is the same in 1986 and 1992: $\frac{P_1^t}{P_2^t} = \frac{P_1}{P_2}$

The table Appendix 1 below gives the detail of calculation for all age categories as it can be done for all age categories, separately for men and women. Except for categories 15-19 and 20-24 where prevalences can be directly derived from observed values, the above strategy applies for all age categories with alternatively 2 or 3 sub-categories involved in the calculation.

Table Appendix 1 here

This approach is replicated for survey years 1981 and 1977 with two further complications:

1. There is no category for younger than 18 in the original data for 1981 and 1977, and five categories only instead of 6 in 1986. Again, we make assumptions to infer the rates based on what is observed: $\frac{P_{19-24}^t}{P_{20-24}^t} = \frac{P_{15-19, 1986}}{P_{20-24, 1986}}$, for $t \in [1977, 1981]$.

2. Using the break down observed in 1992 to produce prevalence rates for 1977 and 1981 is more problematic than when it was used for 1986.
# Tables and graphs

Table 1 – Synoptic of data sources collection

<table>
<thead>
<tr>
<th>Year of production</th>
<th>Population selection</th>
<th>Age range</th>
<th>Number of age groups</th>
<th>Tobacco measurement</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td></td>
<td>15-85</td>
<td>14</td>
<td></td>
<td>27,493</td>
</tr>
<tr>
<td>2005</td>
<td>Random sampling (random generation of phone numbers)</td>
<td>12-75</td>
<td>12</td>
<td>All smokers (regular and occasional)</td>
<td>29,113</td>
</tr>
<tr>
<td>2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13,064</td>
</tr>
<tr>
<td>1995</td>
<td></td>
<td>18-75</td>
<td></td>
<td></td>
<td>1,975</td>
</tr>
<tr>
<td>1992</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2,099</td>
</tr>
<tr>
<td>1986</td>
<td>Method of quotas</td>
<td>15-75</td>
<td>6</td>
<td></td>
<td>2,000</td>
</tr>
<tr>
<td>1981</td>
<td></td>
<td>18-75</td>
<td>5</td>
<td></td>
<td>1,018</td>
</tr>
<tr>
<td>1977</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1,010</td>
</tr>
</tbody>
</table>

*All data produced by INPES (formerly CFES), various poll agencies*
Table 2 – Age categories across versions of surveys

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>15-19</td>
<td>id.</td>
<td>18-19</td>
<td>18-21</td>
<td>15-18</td>
<td>18-24</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
</tr>
<tr>
<td>20-24</td>
<td>id.</td>
<td>22-26</td>
<td>24-26</td>
<td>19-24</td>
<td>18-24</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
</tr>
<tr>
<td>25-29</td>
<td>id.</td>
<td>27-31</td>
<td>26-31</td>
<td>25-34</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
</tr>
<tr>
<td>30-34</td>
<td>id.</td>
<td>32-36</td>
<td>32-36</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
</tr>
<tr>
<td>35-39</td>
<td>id.</td>
<td>37-41</td>
<td>37-41</td>
<td>35-49</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
</tr>
<tr>
<td>40-44</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
<td>42-46</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
</tr>
<tr>
<td>45-49</td>
<td>47-51</td>
<td>47-51</td>
<td>47-51</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
</tr>
<tr>
<td>50-54</td>
<td>id.</td>
<td>52-56</td>
<td>52-56</td>
<td>50-64</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
</tr>
<tr>
<td>55-59</td>
<td>57-61</td>
<td>57-61</td>
<td>57-61</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
</tr>
<tr>
<td>60-64</td>
<td>id.</td>
<td>62-66</td>
<td>62-66</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
</tr>
<tr>
<td>70-74</td>
<td>72-75</td>
<td>72-75</td>
<td>72-75</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
</tr>
<tr>
<td>75-79</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
</tr>
<tr>
<td>80-84</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
<td>id.</td>
</tr>
</tbody>
</table>

Sources: Baromètres Santé (INPES/CFES), CFES surveys.
Table 3 – Proportions of regular among total smokers

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>15-19</td>
<td>0.77</td>
<td>0.76</td>
<td>0.83</td>
<td>0.77</td>
<td>0.70</td>
<td>0.75</td>
<td>0.80</td>
<td>0.87</td>
</tr>
<tr>
<td>20-24</td>
<td>0.82</td>
<td>0.87</td>
<td>0.88</td>
<td>0.91</td>
<td>0.86</td>
<td>0.85</td>
<td>0.87</td>
<td>0.89</td>
</tr>
<tr>
<td>25-29</td>
<td>0.83</td>
<td>0.89</td>
<td>0.88</td>
<td>0.83</td>
<td>0.86</td>
<td>0.84</td>
<td>0.87</td>
<td>0.91</td>
</tr>
<tr>
<td>30-34</td>
<td>0.89</td>
<td>0.88</td>
<td>0.89</td>
<td>0.93</td>
<td>0.89</td>
<td>0.87</td>
<td>0.90</td>
<td>0.85</td>
</tr>
<tr>
<td>35-39</td>
<td>0.89</td>
<td>0.89</td>
<td>0.90</td>
<td>0.98</td>
<td>0.88</td>
<td>0.89</td>
<td>0.87</td>
<td>0.88</td>
</tr>
<tr>
<td>40-44</td>
<td>0.90</td>
<td>0.90</td>
<td>0.85</td>
<td>0.89</td>
<td>0.88</td>
<td>0.86</td>
<td>0.86</td>
<td>0.89</td>
</tr>
<tr>
<td>45-49</td>
<td>0.92</td>
<td>0.89</td>
<td>0.85</td>
<td>0.96</td>
<td>0.89</td>
<td>0.84</td>
<td>0.91</td>
<td>0.96</td>
</tr>
<tr>
<td>50-54</td>
<td>0.91</td>
<td>0.86</td>
<td>0.82</td>
<td>0.95</td>
<td>0.91</td>
<td>0.85</td>
<td>0.79</td>
<td>0.89</td>
</tr>
<tr>
<td>55-59</td>
<td>0.93</td>
<td>0.87</td>
<td>0.87</td>
<td>0.96</td>
<td>0.87</td>
<td>0.85</td>
<td>0.72</td>
<td>0.92</td>
</tr>
<tr>
<td>60-64</td>
<td>0.85</td>
<td>0.87</td>
<td>0.90</td>
<td>0.84</td>
<td>0.87</td>
<td>0.81</td>
<td>0.86</td>
<td>0.64</td>
</tr>
<tr>
<td>65-69</td>
<td>0.78</td>
<td>0.77</td>
<td>0.84</td>
<td>0.80</td>
<td>0.82</td>
<td>0.86</td>
<td>0.87</td>
<td>0.79</td>
</tr>
<tr>
<td>70-74</td>
<td>0.63</td>
<td>0.87</td>
<td>0.89</td>
<td>1.00</td>
<td>0.84</td>
<td>0.90</td>
<td>0.80</td>
<td>1.00</td>
</tr>
<tr>
<td>TOTAL</td>
<td>0.87</td>
<td>0.87</td>
<td>0.87</td>
<td>0.91</td>
<td>0.86</td>
<td>0.85</td>
<td>0.86</td>
<td>0.88</td>
</tr>
</tbody>
</table>

Sources: Baromètres Santé (INPES/CFES), CFES surveys. Method: calculation of authors
Graph 1 – Tobacco prevalence (regular + occasional smokers) by sex and age pooled across generations

Sources: Baromètres Santé (INPES/CFES), CFES surveys. Method: calculation of authors
Table 4 – Tobacco prevalence (regular + occasional smokers) by sex and age across birth cohorts

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1986-90</td>
<td>30.9%</td>
<td>50.8%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1981-85</td>
<td>38.3%</td>
<td>54.0%</td>
<td>58.1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1976-80</td>
<td>32.5%</td>
<td>52.2%</td>
<td>51.5%</td>
<td>53.4%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1971-75</td>
<td>35.9%</td>
<td>55.1%</td>
<td>0.4%</td>
<td>46.5%</td>
<td>46.8%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1966-70</td>
<td>37.0%</td>
<td>58.6%</td>
<td>53.0%</td>
<td>54.0%</td>
<td>44.8%</td>
<td>44.2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1961-65</td>
<td>44.5%</td>
<td>59.0%</td>
<td>61.8%</td>
<td>53.0%</td>
<td>48.6%</td>
<td>38.7%</td>
<td>38.8%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1956-60</td>
<td>40.1%</td>
<td>71.0%</td>
<td>64.7%</td>
<td>52.8%</td>
<td>53.7%</td>
<td>42.8%</td>
<td>38.8%</td>
<td>33.1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1951-55</td>
<td>64.0%</td>
<td>62.6%</td>
<td>55.3%</td>
<td>47.6%</td>
<td>38.9%</td>
<td>34.9%</td>
<td>32.0%</td>
<td>26.8%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1946-50</td>
<td>60.4%</td>
<td>53.4%</td>
<td>51.6%</td>
<td>40.8%</td>
<td>46.5%</td>
<td>30.2%</td>
<td>25.6%</td>
<td>18.6%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1941-45</td>
<td>51.6%</td>
<td>52.7%</td>
<td>46.5%</td>
<td>46.3%</td>
<td>38.8%</td>
<td>30.0%</td>
<td>20.8%</td>
<td>17.9%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1936-40</td>
<td>57.2%</td>
<td>47.5%</td>
<td>39.9%</td>
<td>45.4%</td>
<td>32.0%</td>
<td>20.7%</td>
<td>13.3%</td>
<td>8.1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1931-35</td>
<td>51.6%</td>
<td>40.7%</td>
<td>43.0%</td>
<td>34.2%</td>
<td>18.9%</td>
<td>18.0%</td>
<td>11.1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1926-30</td>
<td>44.2%</td>
<td>49.6%</td>
<td>42.2%</td>
<td>26.4%</td>
<td>19.6%</td>
<td>19.5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1986-90</td>
<td>30.6%</td>
<td>45.2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1981-85</td>
<td>45.7%</td>
<td>43.4%</td>
<td>44.2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1976-80</td>
<td>41.6%</td>
<td>44.6%</td>
<td>38.7%</td>
<td>39.1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1971-75</td>
<td>42.2%</td>
<td>59.4%</td>
<td>44.0%</td>
<td>37.0%</td>
<td>37.1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1966-70</td>
<td>38.0%</td>
<td>52.7%</td>
<td>44.3%</td>
<td>42.2%</td>
<td>39.3%</td>
<td>39.9%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1961-65</td>
<td>36.0%</td>
<td>56.0%</td>
<td>44.7%</td>
<td>35.3%</td>
<td>42.8%</td>
<td>34.8%</td>
<td>33.8%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1956-60</td>
<td>33.9%</td>
<td>53.0%</td>
<td>49.7%</td>
<td>41.6%</td>
<td>39.1%</td>
<td>38.3%</td>
<td>28.7%</td>
<td>31.2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1951-55</td>
<td>50.0%</td>
<td>47.6%</td>
<td>46.3%</td>
<td>41.2%</td>
<td>35.1%</td>
<td>30.3%</td>
<td>22.7%</td>
<td>21.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1946-50</td>
<td>36.2%</td>
<td>44.4%</td>
<td>34.6%</td>
<td>28.9%</td>
<td>23.8%</td>
<td>18.2%</td>
<td>12.5%</td>
<td>14.2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1941-45</td>
<td>33.8%</td>
<td>30.2%</td>
<td>34.3%</td>
<td>13.9%</td>
<td>19.4%</td>
<td>17.3%</td>
<td>8.4%</td>
<td>8.2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1936-40</td>
<td>32.4%</td>
<td>29.9%</td>
<td>24.1%</td>
<td>10.3%</td>
<td>17.5%</td>
<td>11.0%</td>
<td>7.7%</td>
<td>6.2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1931-35</td>
<td>32.1%</td>
<td>21.0%</td>
<td>18.2%</td>
<td>7.8%</td>
<td>8.1%</td>
<td>9.4%</td>
<td>5.5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1926-30</td>
<td>22.5%</td>
<td>20.8%</td>
<td>13.6%</td>
<td>8.5%</td>
<td>12.1%</td>
<td>4.7%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sources: Baromètres Santé (INPES/CFES), CFES surveys. Method: calculation of authors (see §2.3). Cell background: the darker, the cruder the extrapolation.
Graph 2 – Evolution of tobacco prevalence (regular + occasional smokers) by age for different birth cohorts of MALES

Sources: Baromètres Santé (INPES/CFES), CFES surveys. Method: calculation of authors
Graph 3 – Evolution of tobacco prevalence (regular + occasional smokers) by age for different birth cohorts of FEMALES

Sources: Baromètres Santé (INPES/CFES), CFES surveys. Method: calculation of authors
Graph 4 – Evolution of relative prevalence of smoking by age across 6 birth cohorts of MALES

Sources: Baromètres Santé (INPES/CFES), CFES surveys. Method: calculation of authors
Graph 5 – Evolution of relative prevalence of smoking by age across 6 birth cohorts of FEMALES

Sources: Baromètres Santé (INPES/CFES), CFES surveys. Method: calculation of authors
Table 5 – Overview of epidemiological studies assessing the influence of duration/intensity of smoking

<table>
<thead>
<tr>
<th>Source</th>
<th>Scope of the study</th>
<th>Method</th>
<th>Key results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Doll and Peto, 1978</strong></td>
<td>British doctors follow-up, UK</td>
<td>Lung cancer, Males 40-79</td>
<td>Only men continuously smoking/not smoking during the entire follow-up (20 years). Excluded: &gt;40 cigs/day, onset of lung cancer before T0</td>
</tr>
<tr>
<td><strong>Flanders et al., 2003</strong></td>
<td>CPS-II, US</td>
<td>Lung cancer, Both 40-79</td>
<td>Current smokers only Exclusion: &gt;40 cigs/day, onset of lung cancer before T0, started smoking before 10y</td>
</tr>
<tr>
<td><strong>Knoke et al., 2008</strong></td>
<td>CPS-I, US</td>
<td>Lung cancer, Males 40-79</td>
<td>White men. Focus on ex-smokers</td>
</tr>
<tr>
<td><strong>Rachet et al., 2004</strong></td>
<td>Ad hoc case-control cohort, Canada (Quebec)</td>
<td>Lung cancer, Both</td>
<td>Cancer (19 different sites) for case group</td>
</tr>
<tr>
<td><strong>Streppel et al., 2007</strong></td>
<td>Zutphen study (part of the 7 countries study), Netherlands</td>
<td>All causes + 5 specific causes of death (incl. lung cancer), Males 40-80</td>
<td>None</td>
</tr>
<tr>
<td>Age at cessation</td>
<td>2,5</td>
<td>7,5</td>
<td>12,5</td>
</tr>
<tr>
<td>------------------</td>
<td>-----</td>
<td>-----</td>
<td>------</td>
</tr>
<tr>
<td>32</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>37</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>42</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>47</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>52</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>57</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>62</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>67</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>72</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 7 – Key assumptions for the method of reassessment of MAT

<table>
<thead>
<tr>
<th>Assumptions</th>
<th>Default values</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cause of death:</td>
<td>Lung cancer</td>
<td>See discussion of this choice in §4.2.1</td>
</tr>
<tr>
<td>Smoking status:</td>
<td>Overall smoking (regular + occasional)</td>
<td>Restriction to regular smokers difficult to apply (see §5.2 for considerations on feasibility)</td>
</tr>
<tr>
<td>Varying parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration:</td>
<td>Based on long-term prevalence data 1975-2010</td>
<td>As calculated in reference scenario of extrapolation for data (see table 4)</td>
</tr>
<tr>
<td>Cessation:</td>
<td></td>
<td>Alternate scenarios are considered in §5.2</td>
</tr>
<tr>
<td>Age of smoking initiation:</td>
<td>15</td>
<td>Identical for everyone</td>
</tr>
<tr>
<td>Intensity:</td>
<td>20</td>
<td>Might be refined by sex and age substantiated by recent data (ESPS 2010 for example, §5.2)</td>
</tr>
</tbody>
</table>
Table 8 – MAT in 2010 under reference scenario: detailed results

<table>
<thead>
<tr>
<th>Age in 2010</th>
<th>Males</th>
<th></th>
<th>Females</th>
<th></th>
<th>OVERALL</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Current smokers</td>
<td>Ex-smokers</td>
<td>Total</td>
<td>Current smokers</td>
<td>Ex-smokers</td>
<td>Total</td>
</tr>
<tr>
<td>40-44</td>
<td>150</td>
<td>18</td>
<td>168</td>
<td>277</td>
<td>24</td>
<td>301</td>
</tr>
<tr>
<td>45-49</td>
<td>180</td>
<td>39</td>
<td>220</td>
<td>380</td>
<td>70</td>
<td>450</td>
</tr>
<tr>
<td>50-54</td>
<td>1 090</td>
<td>235</td>
<td>1 325</td>
<td>620</td>
<td>133</td>
<td>754</td>
</tr>
<tr>
<td>55-59</td>
<td>1 200</td>
<td>356</td>
<td>1 556</td>
<td>542</td>
<td>221</td>
<td>763</td>
</tr>
<tr>
<td>60-64</td>
<td>1 660</td>
<td>1 180</td>
<td>2 840</td>
<td>711</td>
<td>533</td>
<td>1 243</td>
</tr>
<tr>
<td>65-69</td>
<td>1 341</td>
<td>1 020</td>
<td>2 360</td>
<td>328</td>
<td>544</td>
<td>871</td>
</tr>
<tr>
<td>70-74</td>
<td>945</td>
<td>2 460</td>
<td>3 405</td>
<td>510</td>
<td>1 332</td>
<td>1 841</td>
</tr>
<tr>
<td>75-79</td>
<td>717</td>
<td>3 044</td>
<td>3 761</td>
<td>364</td>
<td>1 413</td>
<td>1 777</td>
</tr>
<tr>
<td>TOTAL</td>
<td>7 282</td>
<td>8 353</td>
<td>15 635</td>
<td>3 733</td>
<td>4 269</td>
<td>8 001</td>
</tr>
</tbody>
</table>

Sources: Baromètres Santé (INPES/CFES), CFES surveys. Method: calculation of authors
Graph 6 – MAT in 2010 (Lung cancer incidence) depending on alternative scenarios on age at initiation and intensity of smoking

Sources: Baromètres Santé (INPES/CFES), CFES surveys. Method: calculation of authors
Table 9 – Observed and predicted tobacco prevalences for several birth cohorts

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>15-19</td>
<td>32.5%</td>
<td>38.3%</td>
<td>30.9%</td>
<td>34.5%</td>
<td>34.6%</td>
</tr>
<tr>
<td>20-24</td>
<td>52.2%</td>
<td>54.0%</td>
<td>50.8%</td>
<td>53.6%</td>
<td>53.1%</td>
</tr>
<tr>
<td>25-29</td>
<td>51.5%</td>
<td>58.1%</td>
<td>50.4%</td>
<td>54.6%</td>
<td>54.6%</td>
</tr>
<tr>
<td>30-34</td>
<td>53.4%</td>
<td>57.7%</td>
<td>49.6%</td>
<td>54.8%</td>
<td>54.3%</td>
</tr>
</tbody>
</table>

*Figures in red italic are predicted values. Greyish cell background spot values for year 2010 (last known prevalences)*
Graph 7 – MAT projection (Lung cancer incidence) in 2010-2060 in reference scenario

Sources: Baromètres Santé (INPES/CFES), CFES surveys. Method: calculation of authors
Table 10 – Examples of simulated prevalences under projection (reference scenario) and simulation (initiation and cessation scenarios)

<table>
<thead>
<tr>
<th>Birth cohort</th>
<th>Year:</th>
<th>2010</th>
<th>2015</th>
<th>2020</th>
<th>2025</th>
<th>2030</th>
<th>2035</th>
<th>2040</th>
<th>2045</th>
<th>2050</th>
<th>2055</th>
<th>2060</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996-00</td>
<td>Age:</td>
<td>15-19</td>
<td>20-24</td>
<td>25-29</td>
<td>30-34</td>
<td>35-39</td>
<td>40-44</td>
<td>45-49</td>
<td>50-54</td>
<td>55-59</td>
<td>60-64</td>
<td></td>
</tr>
<tr>
<td>Projection</td>
<td>Prevalence:</td>
<td>34,6%</td>
<td>53,1%</td>
<td>54,6%</td>
<td>54,3%</td>
<td>50,5%</td>
<td>45,2%</td>
<td>43,0%</td>
<td>36,2%</td>
<td>30,1%</td>
<td>21,1%</td>
<td></td>
</tr>
<tr>
<td>(reference scenario)</td>
<td>Relative cessation rate:</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0,7%</td>
<td>-6,9%</td>
<td>-10,6%</td>
<td>-4,8%</td>
<td>-15,9%</td>
<td>-16,9%</td>
<td>-29,8%</td>
</tr>
<tr>
<td>Simulation:</td>
<td>Prevalence:</td>
<td>17,3%</td>
<td>26,5%</td>
<td>27,3%</td>
<td>27,1%</td>
<td>25,3%</td>
<td>22,6%</td>
<td>21,5%</td>
<td>18,1%</td>
<td>15,0%</td>
<td>10,6%</td>
<td></td>
</tr>
<tr>
<td>-50% initiation</td>
<td>Relative cessation rate:</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-1,0%</td>
<td>-10,4%</td>
<td>-15,8%</td>
<td>-7,2%</td>
<td>-23,9%</td>
<td>-25,3%</td>
<td>-44,7%</td>
</tr>
<tr>
<td>Simulation:</td>
<td>Prevalence:</td>
<td>34,6%</td>
<td>53,1%</td>
<td>54,6%</td>
<td>54,1%</td>
<td>48,5%</td>
<td>40,8%</td>
<td>37,9%</td>
<td>28,8%</td>
<td>21,5%</td>
<td>11,9%</td>
<td></td>
</tr>
<tr>
<td>+50% cessation rate</td>
<td>Relative cessation rate:</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-1,0%</td>
<td>-10,4%</td>
<td>-15,8%</td>
<td>-7,2%</td>
<td>-23,9%</td>
<td>-25,3%</td>
<td>-44,7%</td>
</tr>
<tr>
<td>Projection</td>
<td>Prevalence:</td>
<td>34,5%</td>
<td>53,6%</td>
<td>54,6%</td>
<td>54,8%</td>
<td>50,8%</td>
<td>45,3%</td>
<td>43,2%</td>
<td>36,3%</td>
<td>30,2%</td>
<td>21,2%</td>
<td>17,0%</td>
</tr>
<tr>
<td>(reference scenario)</td>
<td>Relative cessation rate:</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0,0%</td>
<td>-7,2%</td>
<td>-10,9%</td>
<td>-4,6%</td>
<td>-16,1%</td>
<td>-16,8%</td>
<td>-29,8%</td>
</tr>
<tr>
<td>Simulation:</td>
<td>Prevalence:</td>
<td>34,5%</td>
<td>53,6%</td>
<td>54,6%</td>
<td>54,8%</td>
<td>50,8%</td>
<td>45,3%</td>
<td>43,2%</td>
<td>36,3%</td>
<td>30,2%</td>
<td>21,2%</td>
<td>17,0%</td>
</tr>
<tr>
<td>-50% initiation</td>
<td>Relative cessation rate:</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0,0%</td>
<td>-10,9%</td>
<td>-16,3%</td>
<td>-6,8%</td>
<td>-24,1%</td>
<td>-25,3%</td>
<td>-44,7%</td>
</tr>
<tr>
<td>Simulation:</td>
<td>Prevalence:</td>
<td>34,5%</td>
<td>53,6%</td>
<td>54,6%</td>
<td>54,8%</td>
<td>48,8%</td>
<td>40,9%</td>
<td>38,1%</td>
<td>28,9%</td>
<td>21,6%</td>
<td>11,9%</td>
<td>8,4%</td>
</tr>
<tr>
<td>+50% cessation rate</td>
<td>Relative cessation rate:</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0,0%</td>
<td>-10,9%</td>
<td>-16,3%</td>
<td>-6,8%</td>
<td>-24,1%</td>
<td>-25,3%</td>
<td>-44,7%</td>
</tr>
</tbody>
</table>

Sources: Baromètres Santé (INPES/CFES), CFES surveys. Method: calculation of authors
Graph 8 – MAT 2010-2060: reference scenario vs. various scenarios of "reduced initiation"

Sources: Baromètres Santé (INPES/CFES), CFES surveys. Method: calculation of authors
Graph 9 – MAT 2010-2060: reference scenario vs. various scenarios of "increased cessation rate"

Sources: Baromètres Santé (INPES/CFES), CFES surveys. Method: calculation of authors
### Table Appendix1 – Calculations of extrapolated prevalence rates by age for year 1986

<table>
<thead>
<tr>
<th>Age category</th>
<th>Observed prevalence</th>
<th>Aggregate age group</th>
<th>Observed prevalence</th>
<th>Estimated prevalence</th>
<th>Calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td>15-19</td>
<td>$P_1$</td>
<td>15-18</td>
<td>$T_1$</td>
<td>$P'_1 = T_1$</td>
<td></td>
</tr>
<tr>
<td>20-24</td>
<td>$P_2$</td>
<td>19-24</td>
<td>$T_2$</td>
<td>$P'_2 = T_2$</td>
<td></td>
</tr>
<tr>
<td>25-29</td>
<td>$P_3$</td>
<td>25-34</td>
<td>$T_3$</td>
<td>$P'_3 = \frac{2* T_3}{1 + P_4/P_3}$</td>
<td></td>
</tr>
<tr>
<td>30-34</td>
<td>$P_4$</td>
<td>35-49</td>
<td>$T_4$</td>
<td>$P'_4 = \frac{2* T_3}{1 + P_5/P_4}$</td>
<td></td>
</tr>
<tr>
<td>35-39</td>
<td>$P_5$</td>
<td>35-49</td>
<td>$T_4$</td>
<td>$P'_5 = \frac{3* T_4}{1 + P_6/P_5}$</td>
<td></td>
</tr>
<tr>
<td>40-44</td>
<td>$P_6$</td>
<td></td>
<td></td>
<td>$P'_6 = \frac{3* T_4}{1 + P_7/P_6}$</td>
<td></td>
</tr>
<tr>
<td>45-49</td>
<td>$P_7$</td>
<td></td>
<td></td>
<td>$P'_7 = \frac{3* T_4}{1 + P_8/P_7}$</td>
<td></td>
</tr>
<tr>
<td>50-54</td>
<td>$P_8$</td>
<td>50-64</td>
<td>$T_5$</td>
<td>$P'_8 = \frac{3* T_5}{1 + P_9/P_8}$</td>
<td></td>
</tr>
<tr>
<td>55-59</td>
<td>$P_9$</td>
<td></td>
<td></td>
<td>$P'<em>9 = \frac{3* T_5}{1 + P</em>{10}/P_9}$</td>
<td></td>
</tr>
<tr>
<td>60-64</td>
<td>$P_{10}$</td>
<td></td>
<td></td>
<td>$P'<em>{10} = \frac{3* T_5}{1 + P</em>{11}/P_{10}}$</td>
<td></td>
</tr>
<tr>
<td>65-69</td>
<td>$P_{11}$</td>
<td>65+</td>
<td>$T_6$</td>
<td>$P'<em>{11} = \frac{2* T_6}{1 + P</em>{12}/P_{11}}$</td>
<td></td>
</tr>
<tr>
<td>70-74</td>
<td>$P_{12}$</td>
<td></td>
<td></td>
<td>$P'<em>{12} = \frac{2* T_6}{1 + P</em>{13}/P_{12}}$</td>
<td></td>
</tr>
</tbody>
</table>