The impact of parents' health shocks on children's health behaviors

Sylvie Blasco* Eva Moreno - Galbis† Jeremy Tanguy‡§

In progress. Do not cite

Abstract

We evaluate whether parents' health shocks in early childhood, adolescence or adulthood impact their children's risky health behavior. We use a French epidemiological cohort. Two types of health shocks are considered: lung cancer and smoking-related cancer. First, we exploit heterogeneity in the age of the individual at the moment of the parent's health shock to analyze the influence of the cancer diagnosis on the offspring smoking behavior. Second, we propose a Cox proportional hazards model to study the impact of the age of the offspring at the date of the diagnosis on the probability of quitting smoking. Finally, we use the individual smoking history to build a retrospective panel and estimate an individual fixed effects model to identify the impact of the parent's diagnosis on the probability of smoking. In line with the existing literature we find in all cases very limited impact of the parent's health shock on the offspring behavior.

Keywords— smoking, family health shocks, intergenerational transmission of health behaviors
JEL codes— I10, I12, D8, D84

*Le Mans University (France), GAINS-TEPP, IRA, IZA.
†Aix-Marseille Univ., CNRS, EHESS, Centrale Marseille, IRD, AMSE, Marseille, France. Email: eva.moreno-galbis@univ-amu.fr
‡University Savoie Mont Blanc (France), IREGE.
§We thank Constances for providing data access.

1 Introduction

Tobacco is related to a number of diseases and is leading preventable causes of deaths.[1] To minimize these health hazards, active public health policies promoting healthy behaviors, raising awareness about the damages of bad habits, and limiting access to problematic products (via restrictions on consumptions, sales, advertising or via taxes) are implemented worldwide. Still, people do not always respond to these campaigns and a significant share of the population keeps adopting risky health habits. In 2019, 14% of US adults were smokers.These figures call for a better understanding of the determinants of unhealthy behaviors to improve the impact of public health policies. Researches in public health, epidemiology, psychology, sociology and economics have uncovered key determinants of adoption and cessation of risky health behaviors. Great attention has been devoted to intergenerational transmission and to biased assessments of the risks associated with bad habits. In this paper we consider both determinants and investigate whether parents' bad health events related to smoking impact the offspring smoking decisions.

The impact of parents' negative health shock on the offspring's behaviors is a priori undetermined. A parent's diagnosis may act as a shock that helps the individual realizing the actual health hazards of their behavior. It may also alter the individual risk preferences and increase risk aversion (Decker and Schmitz, 2016). Last, it may make the parents adopt healthier behaviors, so that the children evolve in a healthier environment or see their parents start giving a better example. Via these three possible mechanisms, the parents' illness would lower the adoption - and increase the cessation - of risky behaviors. These positive mechanisms could be counterbalanced though, as in face of difficult time (the sickness of a parent) people may on the contrary adopt such behaviors or increase bad habits as a coping mechanism.

We argue here that it is crucial to consider the timing of the parents' diagnosis to fully investigate the issue. The effect of parents' diagnosis on the health behavior is likely to depend on whether the diagnosis is made when the decision of adopting risky behavior has already been made, is about to be made or has not been made yet. The information shock about risks is likely to have greater impact if it happens when the individual is about to initiate smoking. If the diagnosis is made in early childhood, the child will be less exposed to the bad example in case the parent has adopted better habits after their diagnosis. Last, the parents' health shock may deteriorate the socio-economic situation of the offspring and therefore affect their health behaviors differently depending on whether the shock happens during the childhood or at later ages (in the former case, there is a possible risk on investment in education for instance and in the latter case, the individual may have to provide care to the parent).

We use the French Constances data which provides information about the dates and type of parents' diagnosis, along with detailed health outcomes and behaviors of individuals. We look at the timing of the shock, that is whether the parent gets sick during the child's childhood, adolescence or adulthood. We aim at evaluating the impact of smoking-related illness on smoking prevalence and intensity.

To identify causal relations one needs to deal with confounding factors that may explain both the parents' diagnosis and the individual health behavior (preferences, risk perception, socio-economic situation...). We first propose to analyze the link between the parents' diagnosis and the adoption of risky behaviors at the intensive and extensive margins using the rich details of habits described at the time of the interview. Given that we cannot account for the selection into diagnosis due to the lack of information about the individual and family situation at the time of - or before - the diagnosis, we cannot evaluate the impact of having a parent experiencing a health shock versus not having one. In a first step, our approach is to restrict the

[1] Every year around the world 8 million deaths are related to tobacco (WHO, 2018, 2021).

analysis to the sole individuals whose parents experience a health shock and thereby evaluate the impact of the timing of the shock, under the identifying assumption that conditional on having a parent who got sick and conditional on observables, the timing of the parents' health shock is exogenous to the individual. This identifying assumption may be invalid if the timing of the parents' health shock is endogenously determined by the intensity of their risky behaviors. We then propose in a second step to exploit the individual smoking history to build a retrospective panel and estimate an individual fixed effects model having as a dependent variable the probability to smoke.

We find a weak impact of the parents' health shock on the individual risky behavior. Having a parent diagnosed with lung cancer increases the probability to quit smoking if the diagnosis arrives when the offspring is between 15-18 years old. Once we control for individual time invariant heterogeneity we conclude that while the parents' diagnosis does not significantly modify the individual smoking behavior. The effects are even less important if we consider instead other smoking-related cancers. The main determinant of the individual smoking behavior seems rather to be its own diagnosis. In this sense, being diagnosed with a lung cancer strongly reduces the probability of being a smoker while being diagnosed with another smoking-related cancer does not display any significant impact.

To the best of our knowledge, only two studies look at the impact of parents' sickness on the children's behaviors, with a focus on smoking-related illness and smoking decisions. These studies show limited or no impact of parents' health events on the offspring smoking behavior. Darden and Gilleskie (2016) use panel data containing parent and adult offspring data to evaluate whether the smoking-related or general cardiovascular or cancer diagnosis of a parent make the adult offspring change their smoking behavior and their subjective health assessments. They find no effect of such diagnosis for sons, but a significant decrease in both smoking prevalence and intensity conditional on past smoking among daughters following a father's smoking-related cardiovascular event.[2] Li and Gilleskie (2021) use network data to investigate the role of social interactions in smoking behavior. They find no significant effect of parents cardiovascular disease shocks on individual smoking behavior.

Our work is in direct line with the work of Darden and Gilleskie (2016). We extend their work by looking not only at smoking but also at the timing of the parents' health shock. Additionally we do not consider contemporaneous shocks but past shocks. More generally, our paper contributes to the literature that looks at the impact of parents' health shocks on the children's outcomes. The literature has analyzed the impact of parental health shocks (diseases and death) on the health and well-being (Cas et al., 2014; Johnston et al., 2013; Le and Nguyen, 2018), education and skills (Adda et al., 2011; Alam, 2015; Case and Ardington, 2006; Chen et al., 2009; Senne, 2014), cognitive and non-cognitive development and wealth (Arora, 2016) of the children (Le and Nguyen, 2017, 2018) for instance, but not so much on healthy behaviors.

We also contribute to the large literature that investigates the determinants of adopting or quitting unhealthy behaviors. Here the bulk of the research is about intergenerational transmission of behaviors and aims at determining to which extent the observed positive correlation between parents and children's health behavior is the result of a causal impact of parents habits on the children or of confounding factors (see for instance Darden and Gilleskie 2016; Göhlmann et al. 2010; Loureiro et al. 2010; Pan and Han 2017 for smoking).[3]

[2]Hillebrandt (2022) evaluates among others how the death of a partner or a parent affects smoking behavior and body weight. He finds that the death of the father reduces the probability of quitting smoking for daughters, whereas the death of the mother reduces the probability of quitting smoking for both daughters and sons. These results are not in line with those found by Darden and Gilleskie (2016), but note that Hillebrandt (2022) does not restrict the analysis to smoking-related deaths.

[3]See Schmidt and Tauchmann 2011 for alcohol consumption

It is also worth mentioning here papers that look at the impact of own and peer health shocks. Overall the literature finds limited impact of health shocks of social contacts - including spouse, household members, siblings and friends (Sloan et al. 2003 and Khwaja et al. 2006 focus on the response to spouse's health shock; Li and Gilleskie 2021 consider spouse but also friends and family members), but significant decrease in smoking prevalence and intensity following one own diagnosis (Arcidiacono et al., 2007; Bünnings, 2017; Sloan et al., 2003; Smith et al., 2001; Sundmacher, 2012). This larger effect of own health shock may reveal that people need personalized information to update their beliefs about the health risks of bad habits, de facto limiting the effectiveness of general information campaigns.
One could think that the impact could be stronger when the shock happens to parents with whom genetics is shared and identification may be stronger. Our results do not validate this hypothesis. Note though that smokers are not necessarily overly optimistic in their health risk perceptions (Khwaja et al., 2009) and that smokers adjust more drastically their beliefs after a health shock than non smokers or former smokers (Smith et al., 2001), meaning that the higher probability of quitting smoking following a health shock does not necessarily comes from a more accurate belief about health risks. This may explain the limited effect of the parents' diagnosis. Moreover, the health shock needs to be severe enough to make people change their behavior (Darden, 2017), which may explain why we observe that smoking behaviors is more closely related to the diagnosis of lung cancer than to the diagnosis of other types of illness, which may be less perceived as smoke-related.
The paper is organized as follows. Section 2 presents the databases, the variables used as well as descriptive statistics. The econometric strategy is described in section 3. Section 4 presents the estimation results obtained with different econometric approaches as well as robustness tests. Section 5 concludes.

2 Data and descriptive statistics

Constances cohort. We use the Constances data which is a large population-based and general-purpose epidemiological cohort (see Zins et al. (2015) for a detailed description). Started at the end of 2012, it includes a total of about 200,000 individuals. It was designed to be representative of the general French adult population aged between 18 and 69, excluding agricultural and self-employed workers.[4] Participation was voluntary upon invitation sent to randomly selected individuals. In case they agreed to participate in the cohort, individuals were invited by letter to go to one of the 21 Health Screening Centers (HSC), that are partners to the Constances project, to undergo a health examination.[5] Before their health exam, they also had to complete at home two inclusion questionnaires, one about their health, lifestyle and parents' health and other about their complete job history. Participants were then followed annually: they were asked to fill in self-administered follow-up questionnaires once a year, to be returned by postal mail or internet. They finally had to update their health clinic examination every 5 years in a HSC.
All randomly selected individuals invited to participate in the Constances project did not eventually participate and some participants dropped out, which challenges the representativeness of the sample. The

[4]Representativeness was in terms of age and gender.

[5]It is a comprehensive health examination based on standardized operational procedures (SOPs): weight, height, blood pressure, electrocardiogram, vision, auditory, spirometry, and biological parameters; for those aged 45 years and older, a specific work-up of functional, physical, and cognitive capacities are performed. A biobank, including blood and urine samples for each participant is set up. Quality control procedures, including regular on-site visits of research assistants, are delegated to independent organizations.

fact that individuals were randomly drawn from 21 departments is not a threat for representativeness. However, non random acceptation of the invitation and non random attrition are. To deal with non random participation, individuals aged between 18 and 69 living nearby a partner Health Screening Center were drawn randomly by stratified sampling with unequal probabilities, over-representing individuals with a higher probability of non-volunteering according to age.[6]

We compared the age distribution of Constances sample with the age distribution of the sample provided by the French Labor Force Survey (see the Online Appendix). Both the youngest and the oldest age categories are slightly under-represented in Constances, while intermediate age categories are over-represented even if the relative composition of these intermediate categories corresponds well to the LFS.[7]

The socio-demographic composition of the Constances sample is summarized in Table A.1 in Appendix A. As observed, with a share of 53% women are slightly over-represented. Almost 12% of the sample has less than 30 years old, 22% between 30 and 39 years old, 25% between 40 and 49, 21% between 50 and 59 and the rest is above 60 years old. The split across ages is very similar if we consider separately the sample of males and females. 61.4% of the whole sample has at least the Baccalaureate (64% when considering women and 58% for men) and around 26% has completed the Baccalaureate+5 additional years of studies (23% for females and 28% for males). French people stand for more than 92% of the sample and around 69% of the sample is employed. Finally, 59% of our sample is composed of high income people (almost 56% when considering women and 62% for males).

Behavior and health-related outcomes of interest. In this paper we consider smoking behaviors and analyze how the age at which parents are diagnosed a smoking-related diseases influences the smoking behavior of children. The Constances cohort contains rich information about individual health-related outcomes and behaviors. Past health-related behaviors and previous health diagnosis are collected at the first entry into the panel. Information about self-assessed general health, own diagnosis and current health-related behaviors (*e.g.* smoking and alcohol consumption, drug use, dietary habits, sexual activity) is collected at entry and updated annually via the self-administered follow-up questionnaires. As for the parent, we do not have information about their risky behaviors, but we know if they developed a smoking-related illness. This information about the parents' diagnosis is collected at the first interview and is not updated in the follow-up questionnaires. For our econometric analysis we will use data available at the inclusion questionnaires (*i.e.* we do not exploit the longitudinal dimension). A main limitation of the data is that we do not know what was the age of the interviewee when her parents were diagnosed with the disease, but we know the age of the parents at the moment of the diagnosis. We explain below the procedure implemented to approximate the interviewee's age at the moment of the parents' diagnosis.

Table 1 summarizes the main smoking behaviors in our sample. 52% of our sample population has ever

[6] To determine the risk of non-volunteering, Constances used data from participation in previous surveys involving invitations to HSCs. A cohort of non participants was also randomly selected to estimate the probabilities of participation in Constances associated with sociodemographic and health variables and compute weights to correct for non response using administrative data. Unfortunately, this was only done for 2014, so we are unable to use these weights.

[7] Imputing to the Constances sample the average weights computed by gender, nativity and 5 age categories from the French LFS, only leads to an over-representation of the older age category and does not improve the representation of the youngest age category of the Constances sample with respect to the LFS sample (see the Online Appendix). We thus keep the original sample of the Constances survey without implementing any counterfactual reweighting procedure.

smoked in their life, but only 18% are still smoking at the time of the first interview (these proportions increase to 57.5% and 19.6% when for men only).[8] When considering all waves of the panel, among those who ever smoke (panel B), 35% are still smokers at the time of the interview (36.6% for women). They smoke on average 12 cigarettes a day (13.3 for men and 10.6 for women). The average age at which they started smoking was around 17 years old. Those who have quit smoking (panel C) started at a similar age as those who did not (panel B), but their average number of daily cigarettes was slightly higher. They stopped smoking at around 34 years old.[9, 10]

Table 1: Sample composition - Smoking Behaviors

	All	Obs	Males	Obs	Females	Obs
Panel A: Ever smoked and smoking at first wave						
Ever smoked	0.523	96014	0.575	48908	0.478	47106
Smoker	0.185	33923	0.196	16667	0.175	17256
Panel B: Among ever smoked						
Smoker	0.353	33923	0.341	16667	0.366	17256
Age starting smoking	17.438	1648667	17.445	842286	17.430	806381
Number of cigarettes per day	11.975	956992	13.324	543656	10.569	413336
Panel C: Among no longer smokers						
Age start smoking	17.349	1060259	17.318	550930	17.383	509329
Number of cigarettes per day	12.827	657503	14.316	383898	11.194	273605
Age at quitting smoking	34.405	1934303	35.332	1039164	33.388	895139

Concerning parents health, we define for the interviewee's parents several smoking-related illnesses. We consider whether any of the parents was detected a lung cancer or a smoking-related cancer. The former covers not only lung cancer but also bronchial cancer and thoracic cancer. With regard to smoking-related cancers, we follow the French Institute of Cancer and include tongue, throat, nasal, esophageal, laryngeal, otorhinolaryngology, ureter, bladder, breast, uterus, kidney, liver, blood, colon, ovarian, pancreatic, stomach, pharyngeal and bone marrow cancers.

Table 2 shows that 4% of the sample at inclusion have parents that were diagnosed with lung cancer and 21.1% with a smoking-related cancer.[11]

[8] Figure A.1 in Appendix A reveals that the majority of ever smokers started smoking between 15 and 20 years old.

[9] Figure A.2 in Appendix A reveals that most of non-longer smokers stopped smoking between 25 and 45 years old.

[10] Table A.2 in Appendix A reveals that smokers and non-smokers in the first wave have statistically different socio-demographic characteristics. Particularly, the sample of non-smokers is older and is more often composed by already retired people. In the Online Appendix we implement the comparison separating men and women. We still conclude that there are significant differences in the socio-demographic composition of the sample of smokers and non-smokers.

[11] Upward panels of Figure A.5 in Appendix A, reveal that the age at which parents of the interviewee were diagnosed with lung cancer is normally distributed between 35 and 90 years old, with a mode between 55 and 65 years old. For smoking-related cancers, the standard deviation is increased, with the age distribution going from 25 to 90 years old.

Table 2: Proportion of the population with parents having smoking- or alcohol-related illness

	All	Obs	Males	Obs	Females	Obs
Panel A: Smoking-related illness of parents						
Parents with lung cancer	0.040	7411	0.041	3439	0.040	3972
Parents with smoking-related cancer	0.211	38634	0.204	17305	0.217	21329

Table 3 compares the smoking behavior of people depending on whether the parents were diagnosed or not with a cancer. The first three columns of the table focus on lung cancer while columns (4)-(6) consider any smoking-related cancer. From the first three columns we find that the proportion of ever smokers is larger among people whose parents were diagnosed a lung cancer. In contrast, the proportion of current smokers is lower for this population subgroup. Note though that smokers whose parents were diagnosed with a lung cancer smoke a larger quantity of cigarettes per day than smokers whose parents were not diagnosed with cancer. Similar conclusions hold if we focus on the population sample whose parents were diagnosed with a smoking-related cancer (the last three columns of Table 3).[12]

Table 3: Smoking behavior depending on whether parents have cancer or not

	Parents with lung cancer			**Parents with smoking-related cancer**		
	No cancer	Cancer	P-value	No cancer	Cancer	P-value
Ever smoker	0.521	0.572	0.000	0.517	0.544	0.000
Current smoker	0.185	0.168	0.000	0.191	0.162	0.000
Number of cigarettes per day	11.891	13.696	0.000	11.817	12.531	0.000

Computing the age of the interviewee when the parents were diagnosed a smoking- or alcohol-related disease. For the object of our study, one of the limitation of the Constances cohort is that the interviewee only declares the age at which the parent was diagnosed with a smoking-related disease. However, we do not know the age of the interviewee at the moment of the diagnosis. We compute a proxy of this age by exploiting different data sources. Using data from the DADS-EDP panel[13], we compute for all occupations except from people not working (*i.e.* stay-at-home husbands and housewives) the average age at which men and women (separately) have their first child, their second child, the third child and above. The DADS-EDP panel does not provide information on people not working. This is an important limitation since during the second half of the 20th century many women were still housewives. We then use the French

[12]See the Online Appendix for the analysis stratified by gender.

[13]The *Déclarations Annuelles de Données Sociales* (DADS) bring together the mandatory declarations made annually by firms on the wage information of each of their employees for payroll and fiscal purposes. The DADS-EDP panel results from matching the longitudinal version of the DADS, a panel subsample of French salaried employees extracted from exhaustive DADS database for research purposes, with the *Echantillon Démographique Permanent* (EDP) *i.e.* a socio-demographic panel based on the population census since 1967. The DADS panel provides information on employees' characteristics, on their jobs (*e.g.* type of contract, wages and bonuses, number of days worked, hours paid) and on their employers (*e.g.* sector, size, location). The EDP is based on civil status certificates (*e.g.* births, marriages, deaths) and covers since 2002 individuals born on October 1-4, on January 2-5, on April 1-4 and on July 1-4.

Labor Force Survey 1993 to compute a proxy of the age at which people not working had their first child. Information in the LFS is less detailed than in the DADS-EDP panel. We do not have an explicit question on the first child. We do know though if the individual has a child below 3 years old. We then consider the population sample declaring having a child below 3 and for whom the number of children below 3 equals the number of children below 18. In this way we are almost sure that we are isolating the first child of the men or women[14].

We implement the following procedure. First, we compute the average age of the individual not working and declaring having a child below 3 years old. We subtract 3 years from this average age, so as to compute a lower bound of the age of the first child. Second, we compute by gender the average age in 1993 of the first child for the whole population using the same procedure as for not working people. Third, we compute the standard deviation between the age of the first child for not working individuals and the average age of the first child for the whole population by gender. Fourth, we use DADS-EDP panel and compute for every year between 1968 and 2000 the average age of the first child by gender and apply the standard deviation computed for not working individuals in 1993. This allows us to compute the approximate age of the first child for not working men and women for the 1968-2000 period. Note that for not working people we will only have information about the age of the first child, while when considering working people (data coming from the DADS-EDP panel) we have information on the age of the first child, second child and third (or more) child.

Since Constances provides information on the occupation of the interviewee's parents and on the rank of the interviewee within her siblings, we can impute for each interviewee the approximative age of her parents at the time of her birth. Then, given that we have information on the age at which the parents were diagnosed with a smoking-related disease, we can estimate the approximate age of the individual when the parents were diagnosed with the disease. Note that for the object of our study it is not very important if the age of the interviewee at the moment of the diagnosis of the parents disease is approximative, since our focus will be on age categories and not on the exact age. That is, we will group together individuals that were below 15 years old at the moment of the diagnosis, between 15 and 18, between 18 and 25 etc. Moreover, we will propose as a robustness test to use broader age categories.

3 Empirical strategy

We aim at evaluating the impact of a parents' health shock on the offspring's smoking behavior. To identify the impact of a parents' health shock, one need to control for the parents' health-related behaviors which are correlated with both the parents' diagnosis and the individual health-related behaviors. Indeed, the probability of getting sick and being diagnosed is greater for parents adopting risky behaviors. Moreover, due to the intergenerational transmission of behaviors, individuals are more likely to adopt the same habits as their parents. Put differently, individuals with and without a parent diagnosed are most likely to differ with respect to family backgrounds and parents' health behavior. Not accounting for this selection into the parents' diagnosis could severely downward bias our estimate of the impact of the health shock, as the possibly limiting effect of the diagnosis on risky behavior would be counteracted by the fact that those more likely to adopt unhealthy behaviors are also those more likely to have a parent who got diagnosed.

[14] Evidently, we could have a situation where there is an age difference within the fraternity above 15 years, which will lead us to consider as a first child a kid that is actually not the first one for the father and/or mother. We believe though that this situation remains unfrequent.

Unfortunately, the data at hand do not provide information about the smoking behaviors of the parents or information about the individual at the time of the diagnosis.

To remedy this issue, we proceed in two steps. First, we restrict the analysis sample to individuals whose parents were diagnosed with a smoking-related disease. We only include in our sample individuals whose parents have been diagnosed with a lung cancer or any other smoking-related cancer. Doing so, we perform the analysis on a more homogeneous sample, limiting the selection bias: because the majority of these parents are likely to have engaged in risky smoking behaviors, we manage to cancel out, at least partially, the effect on the individual coming from the parents' behavior. This strategy implies that we evaluate the impact of the timing of the diagnosis (given that a parent got diagnosed), instead of the impact of having a parent who got diagnosed. Our identifying assumption is that conditional on having a parent diagnosed, the timing of the diagnosis is exogenous to the individual. This identifying assumption may be questionable, since an individual can influence the timing of the diagnosis by intensifying her risky behavior. We therefore propose an alternative approach that exploits the individual smoking history to build a retrospective panel and allows to implement an estimation controlling for individual fixed effects.

First approach: evaluation of the effect of the timing of the diagnosis. We first estimate the following type of econometric model on a sample of individuals whose parents where diagnosed with a smoking-related illness:

$$
\begin{aligned}
y_i \quad = \quad & \beta_0 + \delta_{<15} ParentDiag_{<15} + \delta_{15-18} ParentDiag_{15-18} + \delta_{19-24} ParentDiag_{19-24} \\
& + X_i\beta_1 + \beta_2 OwnDiag_i + \gamma_t + \gamma_t^2 + u_i
\end{aligned}
$$

Where the δ_k coefficient measures the effect of having a parent's health shock at age $k = \{< 15; 15 - 18; 19 - 24\}$ instead of having the shock after 25. We control for a number of individual attributes, including whether the individual was diagnosed herself at some point before the time of the interview with lung cancer or smoking-related cancer. We also control for the individual's mental health,[15] and her subjective perception of her general health status. Additionally, we include standard socio-demographic variables (gender, age, cohabitation status, place of birth, education, employment situation and socio-professional category). We also control for time trend in a flexible form to account for the action of changing awareness and cultural norms vis-à-vis smoking along time (due to public health policies for instance).[16]

We look at whether the individual has ever been a smoker, at the age at which the individual starts smoking and at the average number of cigarettes per day and evaluate the effect of the age at which one of the parents was diagnosed with lung cancer or with another smoking-related cancer..

Given that we cannot control for the parents smoking and drinking behavior, the parameters of interest, δs, capture both the effect of the parents' health-related behavior and the effect of the parents' diagnosis. The former implies that parents who smoke on a regular basis will transmit to their children this type of behavior. Children are then likely to replicate this risky behavior. In contrast, the impact of a parents'

[15]Our mental health indicator is the CES-D scale, that measures the mood and mental state of individuals in the past week. It includes 20 items that cover the main symptoms of depression. Questions allowing to construct the CES-D scale are asked at inclusion. Participants answer each item on a four-point scale on which they indicate the frequency with which they experienced the corresponding symptom during the past week: 0 = rarely or none of the time (less than 1 day); 1 = some or little of the time (1-2 days); 2 = occasionally or a moderate amount of the time (3-4 days); and 3 = most or all of the time (5-7 days). The total score ranges from 0 to 60, with higher scores indicating more symptoms and therefore worse mental health. See Morin et al. (2011) for further details.

[16]We include a linear and a non linear trend on the date of inclusion as well as a trend capturing the date of birth.

diagnosis is likely to be the opposite and discourage children to replicate the parents' behavior so as to avoid a replication of the parent's health outcome. Our guess is that the intensity of both contradictory effects will differ along the children life-cycle. More precisely, an early diagnosis may reflect the fact that the parents were heavy smokers. According to the behavior transmission story, this should have increased the probability of being a smoker of the children. However, teenagers are less likely to be addicted since they have not been exposed during longtime to the behavior. Moreover, if the diagnosis makes the parent quit the unhealthy behavior or helps the individual realize the health hazards associated with smoking, we would expect the diagnosis to have a negative impact on the probability of smoking or drinking of the children. Therefore, our guess is that, when children are young, the negative impact of a parent's diagnosis is likely to dominate over the impact of the transmission behavior since offsprings have been exposed to the unhealthy behavior for a limited period of time.

The older the children at the moment of the parent's diagnosis the more likely is the transmission behavior effect to dominate over the parent's diagnosis effect. If an individual has been smoking for more than 10 years, she is likely to be already addicted. The arrival of a negative shock coming from a parent's diagnosis is less likely to modify the offspring unhealthy behavior since the addiction mechanism is at play. Moreover, a bad parent's diagnosis may increase anxiety and stress of the offspring which may push the individual to smoke even more.

Second approach: a fixed effects model to relax the identifying assumption about the timing of the diagnosis. The timing of the diagnosis may not be exogenous to the individual in case individuals with riskier behaviors or with specific life conditions are more likely to be diagnosed earlier than others. In that case, there may be confounding factors and individuals whose parents are diagnosed at earlier ages are not comparable to those whose parents are diagnosed later. To remedy this issue, we use our data as a retrospective panel and estimate a fixed effects model. Individual fixed effects control for the individual time-invariant unobserved heterogeneity, such as time preference, family situation in the childhood and parents health behaviors in the childhood. Given that we have the age of first cigarette, in case the individual is a smoker, and the age of the last cigarette (in case the individual quit smoking), we construct the individual's smoking history and we are able to know at each age whether the individual is a smoker or not and whether her parents had already been diagnosed or not. Note that we do not have retrospective information about the number of cigarettes smoked per day for each year, so we cannot perform this fixed effects analysis at the intensive margins.

We therefore estimate the following linear probability model with individual fixed effects:

$$y_{it} = \beta_0 + \delta PostParentDiag_{it} + \beta_2 PostOwnDiag_{it} + \beta_3 AgeDummies + \beta_4 YearDummies + \alpha_i + \epsilon_{it}$$

where y_{it} equals unity if individual i is a smoker at date t. The variable $PostParentDiag_{it}$ equals zero if none of the individual's parents has ever been diagnosed with cancer at date t and equals unity if at least one of the individual's parents has been diagnosed with cancer in the past. Actually, this variable equals unity from the year following the parent's diagnosis until present. The variable $PostOwnDiag_{it}$ is defined in the same way but concerns the interviewee's own diagnosis. $AgeDummies$ will be the different age categories we consider. We also control for year fixed effects ($YearDummies$) and for individual fixed effects (α_i).

4 Results

4.1 Benchmark estimations

The impact of the age at which parents were diagnosed with lung cancer or with another smoking-related cancer on the offspring smoking behavior is summarized, respectively, in Panels A and B in Table 4. As observed from Panel A having at least one parent diagnosed with lung cancer when the individual is below 25 years old does not significantly modify the fact of ever smoking, the age at which the individual started smoking or the number of cigarettes per day with respect to receiving the diagnosis when the individual is above 25.[17]

Table 4: Influence of the age at which the individual's parents were diagnosed with lung cancer on the individual's smoking behavior.

	(1) Ever smoked	(2) Age start smoking	(3) Number of cigarettes per day
Panel A: Lung cancer			
Under 15 years old	-0.005	-0.423	0.504
	(0.043)	(0.304)	(0.972)
15-18 years old	-0.016	-0.075	-0.238
	(0.039)	(0.354)	(0.821)
19-24 years old	0.007	-0.158	0.647
	(0.024)	(0.239)	(0.563)
Observations	4752	2675	2436
R-squared	0.044	0.059	0.093
Panel B: Smoking-related cancer			
Under 15 years old	0.027**	-0.257**	0.953***
	(0.013)	(0.116)	(0.329)
15-18 years old	0.014	0.113	0.510*
	(0.014)	(0.155)	(0.305)
19-24 years old	0.028***	-0.088	0.420*
	(0.010)	(0.108)	(0.223)
Observations	25267	13420	11825
R-squared	0.039	0.041	0.102
Trends (time, time2, occupation)	Yes	Yes	Yes
Department FE	Yes	Yes	Yes
Demographics	Yes	Yes	Yes
Educational attainment	Yes	Yes	Yes
Employment-related controls	Yes	Yes	Yes
Health-related controls	Yes	Yes	Yes
Work-related controls	No	No	No
Sample	All	All	All

Source: Constances (2012-2018). Notes: Robust standard errors in parentheses. Statistical significance: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

Panel B reveals that when considering the "Ever smoked" indicator we conclude that individuals whose parents were diagnosed with a smoking-related cancer when they were below 15 years old or between 19

[17] The analysis by gender confirms this finding. See Online Appendix

and 24 years old, are more likely to have ever smoked in their life with respect to individuals whose parents were diagnosed with a smoking-related cancer when they were above 24 years old. This suggests that the individual may not be aware that smoking is related to the type of cancer diagnosed to their parents.
This hypothesis is confirmed when we focus on the age at which the individual starts smoking. As we see, from Panel B, people who were below 15 years old when their parents were diagnosed started smoking significantly earlier than people who were above 24 at the moment of the diagnosis. In the same line, the number of cigarettes smoked per day by people whose parents were diagnosed when they were below 15 years old is significantly larger than for people who were above 24 at the moment of the diagnosis.[18] Again, by means of their behavior, individuals reveal that they are not associating their parents' cancer with smoking, so their risky behavior is not influenced by the timing of the diagnosis.

4.1.1 Cohort effects

Estimations in the previous section included a linear trend on the age of birth. We now replace this trend by allowing instead the impact of the age at the moment of the diagnosis to vary across cohorts. The intuition is that the individual's perception of the health risk associated with smoking behavior is likely to have evolved along time, as scientists have managed to prove causal relationships between these types of behaviors and cancers or other diseases. This change in perception is then likely to be cohort-specific, that is strongly related to the date of birth of the individual. We proceed as follows. First, we compute the age distribution over the whole sample. Second, we define four quartiles of the age distribution. Finally, we consider again the sample of people whose parents were diagnosed with lung cancer or smoking-related cancer and we replicate the benchmark regression but interacting the age quartile to which the interviewed individual belongs with the age she had at the moment where her parents were diagnosed. With this approach, we are able to distinguish the differentiated impact of an identical age of parents' diagnosis on people belonging to different generations. For instance, someone who is currently 60 years old and whose parents were diagnosed with lung cancer when she was below 15, may not have the same reaction than someone who is currently 30 and whose parents were diagnosed with lung cancer when she was below 15.
Estimation results for the sample of people whose parents were diagnosed with lung cancer are summarized in Table 5. The reference category is people located on the fourth quartile of the age distribution, that is, the oldest people of the sample. The first column of the table reveals that people with a parent diagnosed with lung cancer when they were below 15 years old, are less likely to ever smoke if they belong to the first or the third quartile of the age distribution, with respect to people who were the same age at the moment of the diagnosis but are in the fourth quartile of the age distribution. The same applies for people who were 15-18 years old at the moment of the diagnosis and belong to the third quartile. [19] For identical age at the moment of the parents' diagnosis, younger generations are more likely to decrease their risky behavior than older generations, suggesting that younger people have already updated the proven scientific link between tobacco and cancer.

[18]Coefficients associated with 15-18 and 19-24 years old are positive but only significant at 10%. See Online Appendix for an analysis by gender.

[19]The analysis by gender is available in the Online Appendix.

Table 5: Influence of the age at which the individual's parents were diagnosed with lung cancer on the individual's smoking behavior. Cohort effects.

	(1)	(2)	(3)
	Ever smoked	Age start smoking	Number of cigarettes per day
Under 15 years old	0.241**	-0.976	-0.313
	(0.112)	(0.611)	(2.407)
15-18 years old	0.165	-0.741	1.630
	(0.129)	(0.469)	(2.883)
19-24 years old	-0.007	-0.303	2.946*
	(0.056)	(0.479)	(1.698)
Age quartile 1	-0.025	1.280*	-0.242
	(0.077)	(0.721)	(1.738)
Age quartile 1 · Under 15 years old	-0.352***	1.099	0.125
	(0.135)	(0.820)	(2.785)
Age quartile 1 · 15-18 years old	-0.114	-0.054	-2.173
	(0.146)	(0.675)	(3.116)
Age quartile 1 · 19-24 years old	0.062	0.169	-2.489
	(0.076)	(0.634)	(2.031)
Age quartile 2	-0.045	1.225**	-1.133
	(0.056)	(0.572)	(1.428)
Age quartile 2 · Under 15 years old	-0.181	1.158	0.687
	(0.133)	(0.805)	(2.875)
Age quartile 2 · 15-18 years old	-0.186	1.085	-2.062
	(0.144)	(0.805)	(3.246)
Age quartile 2 · 19-24 years old	-0.024	0.633	-3.694*
	(0.069)	(0.713)	(1.884)
Age quartile 3	-0.013	0.399	-0.421
	(0.038)	(0.406)	(1.041)
Age quartile 3 · Under 15 years old	-0.282*	-0.913	2.027
	(0.144)	(0.837)	(3.669)
Age quartile 3 · 15-18 years old	-0.317**	1.512	-2.456
	(0.149)	(1.018)	(3.431)
Age quartile 3 · 19-24 years old	0.029	-0.277	-1.976
	(0.073)	(0.607)	(2.087)
Trends (time, time2, occupation)	Yes	Yes	Yes
Department FE	Yes	Yes	Yes
Demographics	Yes	Yes	Yes
Educational attainment	Yes	Yes	Yes
Employment-related controls	Yes	Yes	Yes
Work-related controls	No	No	No
Health-related controls	Yes	Yes	Yes
Sample	All	All	All
Observations	4752	2675	2436
R-squared	0.049	0.065	0.099

Source: Constances (2012-2018). Notes: Robust standard errors in parentheses. Statistical significance: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

When considering the age of the first cigarette, we simply conclude that people in the first and second quartile of the distribution started smoking significantly latter than older generations. For the average number of cigarettes per day no coefficient is significant at 5%.[20]
Results are less interesting when considering the sample of people whose parents were diagnosed with any smoking-related cancer (see Table 6). In this case we can only conclude that younger generations (belonging to the second quartile of the age distribution) are less likely to ever smoke and delay the age of the first cigarette with respect to the oldest cohort independently on the age at which the parents were diagnosed with cancer.[21] This confirms that the age of the parents diagnosis has no significant role in the behavior.[22]

[20]When considering 10% as the threshold of significance, we find that people whose parents were diagnosed with lung cancer when they were 19-24 years old consume significantly more cigarettes than those who were above 24 at the moment of the diagnosis unless they belong to the third quartile of the age distribution.

[21]The analysis by gender is proposed in Tables **??** and **??** in Appendix **??**

[22]We propose in the Online Appendix a robustness check where the age categories at the moment of the parents' diagnosis are larger: less than 15 years old, between 15 and 24 years old and 25 years old or more. Results are perfectly in line with findings from Table 4 or with cohorts analyses.

Table 6: Influence of the age at which the individual's parents were diagnosed with a smoking-related cancer on the individual's smoking behavior. Cohort effects.

	(1)	(2)	(3)
	Ever smoked	Age start smoking	Number of cigarettes per day
Under 15 years old	-0.014	-0.199	1.833
	(0.042)	(0.563)	(1.448)
15-18 years old	0.018	0.748	0.112
	(0.036)	(0.472)	(1.003)
19-24 years old	0.013	0.277	0.309
	(0.026)	(0.403)	(0.753)
Age quartile 1	-0.054	0.394	-0.192
	(0.033)	(0.340)	(0.744)
Age quartile 1 · Under 15 years old	0.027	-0.072	-1.540
	(0.048)	(0.590)	(1.508)
Age quartile 1 · 15-18 years old	-0.014	-0.908*	0.606
	(0.044)	(0.521)	(1.122)
Age quartile 1 · 19-24 years old	0.005	-0.369	0.220
	(0.032)	(0.440)	(0.835)
Age quartile 2	-0.069***	0.673**	-0.325
	(0.024)	(0.273)	(0.595)
Age quartile 2 · Under 15 years old	0.081*	-0.146	-0.907
	(0.048)	(0.590)	(1.536)
Age quartile 2 · 15-18 years old	-0.021	-0.596	0.422
	(0.044)	(0.560)	(1.142)
Age quartile 2 · 19-24 years old	0.034	-0.394	0.268
	(0.031)	(0.435)	(0.825)
Age quartile 3	-0.027	0.276	-0.438
	(0.017)	(0.190)	(0.425)
Age quartile 3 · Under 15 years old	0.011	0.048	-0.067
	(0.052)	(0.622)	(1.713)
Age quartile 3 · 15-18 years old	0.016	-0.627	0.107
	(0.045)	(0.573)	(1.166)
Age quartile 3 · 19-24 years old	0.000	-0.611	-0.303
	(0.033)	(0.457)	(0.891)
Trends (time, time2, occupation)	Yes	Yes	Yes
Department FE	Yes	Yes	Yes
Demographics	Yes	Yes	Yes
Educational attainment	Yes	Yes	Yes
Employment-related controls	Yes	Yes	Yes
Work-related controls	No	No	No
Health-related controls	Yes	Yes	Yes
Sample	All	All	All
Observations	25267	13420	11825
R-squared	0.041	0.042	0.108

Source: Constances (2012-2018). Notes: Robust standard errors in parentheses. Statistical significance: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

4.2 How smoking duration is influenced by the age at which the parents were diagnosed with cancer?

In the previous section we have analyzed how differences in the age at which the individual's parents were diagnosed with cancer influence the individual's smoking behavior. Since Constances also provides for ever smokers the date at which the individual quit smoking (if she has quit), we analyze the survival time from the start to the end of smoking to investigate how the age at which the individual parents were diagnosed with cancer affects the probability of quitting smoking. We use a non-parametric Cox model based on the assumption of proportional hazards.

Results are displayed in Table 7. In column (1), no control variable or fixed effect is introduced. In this case we find that there is a positive and significant effect on the probability of quitting smoking if one of parents was diagnosed with a lung cancer when the individual was between 15-18 years old and between 25-29 years old (the reference being more than 49 years old).[23] Column (2) adds fixed effects for department of residence, year of the first cigarette, age of the first cigarette. Demographic and educational controls are added in columns (3) and (4). Estimates in column (3) are based on the proportional hazards assumption, while estimates in column (4) allow non-proportional hazards. Results are almost identical, suggesting that assuming proportional hazards is a reasonable choice. From both columns we conclude that having a parent diagnosed with lung cancer when the individual is between 15-18 years old significantly increases the probability of quitting smoking with respect to the situation where the individual is above 49 years old at the moment of the diagnosis.[24]

[23] For people between 40-49 years old the coefficient is significant at 10%.

[24] Again for people between 40-49 years old the coefficient is significant at 10%.

Table 7: Influence of the age at which the individual's parents were diagnosed with lung cancer on the individual's smoking duration: all smokers

Hazard ratios	(1)	(2)	(3)	(4)
Under 15 years old	1.205	1.079	1.046	1.050
	(0.187)	(0.175)	(0.173)	(0.174)
15-18 years old	1.783***	1.514***	1.720***	1.714***
	(0.248)	(0.223)	(0.258)	(0.258)
19-24 years old	1.057	0.946	1.023	1.024
	(0.106)	(0.100)	(0.113)	(0.113)
25-29 years old	1.229**	1.107	1.152	1.148
	(0.116)	(0.110)	(0.119)	(0.119)
30-39 years old	1.104	1.018	1.078	1.073
	(0.083)	(0.081)	(0.089)	(0.089)
40-49 years old	1.141*	1.105	1.152*	1.152*
	(0.087)	(0.087)	(0.094)	(0.094)
Smoking starting year FE	No	Yes	Yes	Yes
Smoking starting age FE	No	Yes	Yes	Yes
Department FE	No	Yes	Yes	Yes
Demographics	No	No	Yes	Yes
Educational attainment	No	No	Yes	Yes
Employment-related controls	No	No	No	No
Health-related controls	No	No	No	No
Work-related controls	No	No	No	No
Sample	Smokers	Smokers	Smokers	Smokers
Model	CoxPH	CoxPH	CoxPH	CoxNPH
Observations	2968	2968	2803	61375

Source: Constances (2012-2018). Notes: Robust standard errors in parentheses. Statistical significance: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

Table 8 considers different samples. Column (1) focuses on smokers who started smoking before the diagnosis and quit smoking after the diagnosis. Unsurprisingly for this sample, coefficients associated with every age category are positive and very significant, since we are imposing a sample of people that quit smoking after the parents diagnosis (the reference age is people above 49 years old). Column (2) considers smokers who quit smoking after the diagnosis but they could have started smoking before or after the parents' diagnosis. Results from column (2) are similar to those in column (1) since we are focusing on people who quit smoking after the parents receive the diagnosis of lung cancer. Column (3) is the least constrained sample, since it focuses on people who started smoking before the diagnosis but who might currently be smokers (they did not necessarily quit smoking after the parents diagnosis). In that case, we find that only people whose parents were diagnosed with lung cancer when they were between 15-18 years old have a significantly larger probability of quitting smoking than people whose parents were diagnosed when they were above 49 years old.[25] This sub-sample seems to be the main driver of the results estimated over the whole sample of individuals in Table 7.

[25]The coefficient associated with individuals between 40-49 years old is significant at 10%.

Table 8: Influence of the age at which the individual's parents were diagnosed with lung cancer on the individual's smoking duration: comparison between samples of smokers

Hazard ratios	(1)	(2)	(3)
Under 15 years old	10.711***	8.379***	1.105
	(8.435)	(2.351)	(0.804)
15-18 years old	16.666***	15.176***	1.689***
	(4.878)	(4.197)	(0.291)
19-24 years old	7.427***	7.390***	1.010
	(1.890)	(1.874)	(0.113)
25-29 years old	6.956***	6.568***	1.143
	(1.762)	(1.668)	(0.117)
30-39 years old	4.259***	4.109***	1.068
	(1.036)	(1.004)	(0.087)
40-49 years old	2.783***	2.784***	1.143*
	(0.691)	(0.695)	(0.092)
Smoking starting year FE	Yes	Yes	Yes
Smoking starting age FE	Yes	Yes	Yes
Department FE	Yes	Yes	Yes
Demographics	Yes	Yes	Yes
Educational attainment	Yes	Yes	Yes
Employment-related controls	No	No	No
Health-related controls	No	No	No
Work-related controls	No	No	No
Sample	SmokersA	SmokersB	SmokersC
Model	CoxPH	CoxPH	CoxPH
Observations	1494	1642	2655

Source: Constances (2012-2018). Notes: Robust standard errors in parentheses. Statistical significance: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. SmokersA are smokers who started smoking before the diagnosis and quit smoking after the diagnosis. SmokersB are smokers who quit smoking after the diagnosis. SmokersC are smokers who started smoking before the diagnosis.

Table B.1 in Appendix B replicates the same estimations as in Table 7 but considering any smoking-related cancer other than lung cancer. Consistently with previous findings, the age of the parents' diagnosis of a smoking-related cancer does not significantly influence the probability of quitting smoking whether we assume proportional hazards or not. Table 9 splits again our sample in the same three sub-samples considered in Table 8. Column (1) considers only smokers who started smoking before the diagnosis and quit smoking after the diagnosis. Again, the age of the diagnosis positively and significantly influences the probability of quitting smoking. Column (2) covers smokers who quit smoking after the diagnosis but that may have started smoking before or after the diagnosis. The probability of quitting smoking is still significantly increased whatever the age of the diagnosis we consider. Finally, as soon as we consider the sample of smokers who started smoking before the diagnosis but who may still be smokers or not, results are modified. Now, no significant coefficient arises suggesting that having a parent diagnosed with a smoking-related cancer does not affect the smoking behavior of the individual. Furthermore, these people are the main drivers of the result obtained over the aggregate sample in Table B.1.

Table 9: Influence of the age at which the individual's parents were diagnosed with smoking-related cancer on the individual's smoking duration: comparison between samples of smokers

Hazard ratios	(1)	(2)	(3)
Under 15 years old	6.301***	6.781***	0.802
	(1.625)	(0.675)	(0.192)
15-18 years old	8.552***	7.056***	1.087
	(0.938)	(0.701)	(0.071)
19-24 years old	6.808***	6.292***	0.999
	(0.663)	(0.593)	(0.042)
25-29 years old	5.585***	5.075***	1.030
	(0.538)	(0.476)	(0.040)
30-39 years old	3.709***	3.465***	0.988
	(0.347)	(0.316)	(0.032)
40-49 years old	2.290***	2.190***	1.019
	(0.222)	(0.208)	(0.033)
Smoking starting year FE	Yes	Yes	Yes
Smoking starting age FE	Yes	Yes	Yes
Department FE	Yes	Yes	Yes
Demographics	Yes	Yes	Yes
Educational attainment	Yes	Yes	Yes
Employment-related controls	No	No	No
Health-related controls	No	No	No
Work-related controls	No	No	No
Sample	SmokersA	SmokersB	SmokersC
Model	CoxPH	CoxPH	CoxPH
Observations	8507	9988	14924

Source: Constances (2012-2018). Notes: Robust standard errors in parentheses. Statistical significance: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. SmokersA are smokers who started smoking before the diagnosis and quit smoking after the diagnosis. SmokersB are smokers who quit smoking after the diagnosis. SmokersC are smokers who started smoking before the diagnosis.

4.3 Estimation results from a fixed effects model

As remarked in the previous section, the main advantage of a fixed effects approach is that it allows us to circumvent potential criticisms associated with the endogeneity of the date of the diagnosis. The construction of a retrospective panel allows for the inclusion of individual fixed effects to control for the individual time-invariant unobserved heterogeneity. In this way unobserved differences in the intensity of parents' risky behaviors, which may explain differences in date of diagnosis, are absorbed.

Our benchmark results of the fixed effects model are reported in Table 10. The dependent variable, "Smoker", equals unity for every age of the individual during which she declares to be a smoker. Columns (1) and (3) control for the age of the individual by considering 4 age categories: under 15, between 15 and 18 and between 19 and 24, the reference category is over 24. Columns (2) and (4) consider a more detailed definition of age categories. less than 15, 15-18, 19-24, 25-29, 30-39, 40-49, 50-59, and 60 or more. The reference category in this case is people between 30 and 39, since it is the most numerous age category.

The fact of having a parent diagnosed with a cancer is captured by the variable "Parents' diagnosis" which equals unity for every age of the individual following the date of the parents' diagnosis. Columns (1)-(2) focus on lung cancer, while columns (3)-(4) consider other smoking-related cancers. Finally, we also introduce as explanatory variable the individual's own diagnosis, of lung cancer or any other smoking-related cancer.

Columns (1) and (3), results in Table 10 reveal that having a parent diagnosed with cancer reduces the probability of being a smoker, whether lung cancer or any other smoking-related cancer. However coefficients are only significant at 10%. The same two columns show that individuals having been diagnosed themselves with lung cancer are significantly less likely to be smokers. In contrast, we do not find any significant effect of own diagnosis for other smoking-related cancers.

Columns (2) and (4) control for a larger set of age categories. We still find a negative, but not significant, coefficient of having a parent diagnosed with lung cancer. For other smoking-related cancers, the impact of the diagnosis is zero and not significant. The effect of own diagnosis remains negative and significant for lung cancer but is not significant for other smoking-related cancers.

all in all, given the non-significance (or weak significance) of coefficients associated with parents' diagnosis we cannot conclude that a health shock to the parents modifies the offspring behavior.

Table 10: Average effect of parents' cancer diagnosis on smoking

	(1)	(2)	(3)	(4)
	Lung cancer	Lung cancer	Smoking-related cancer	Smoking-related cancer
Parents' diagnosis	-0.009*	-0.004	-0.004*	0.000
	(0.005)	(0.005)	(0.002)	(0.002)
Own diagnosis	-0.193***	-0.201***	0.003	-0.000
	(0.035)	(0.035)	(0.005)	(0.005)
Intercept	0.658***	0.467***	0.659***	0.467***
	(0.022)	(0.022)	(0.022)	(0.022)
Dependent variable	Smoker	Smoker	Smoker	Smoker
Year fixed effects	Yes	Yes	Yes	Yes
Individual fixed effects	Yes	Yes	Yes	Yes
4 age categories	Yes	No	Yes	No
8 age categories	No	Yes	No	Yes
Sample	All	All	All	All
Observations	200312	200312	200312	200312
R-squared	0.633	0.633	0.633	0.633

Source: Constances (2012-2019). Notes: Robust standard errors are in parentheses and clustered at the individual level. Statistical significance: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

4.3.1 Estimations results from a fixed effects model: an analysis by age categories

In this section we implement separate regressions by splitting the sample depending on the individual's age at the time of parent diagnosis. More precisely, we define j samples with $j = 1, 2, 3, 4, 5, 6, 7$ corresponding to the age category at the time of parent diagnosis. Every sample j is composed by all individuals whose parents were never diagnosed with cancer plus individuals whose parents were diagnosed when the individual was in the age category j. In this case, the coefficient associated with the variable "Parents' diagnosis" measures the difference in the likelihood of smoking between individuals with a diagnosed parent and those with undiagnosed parents.

Table 11 covers all individuals whose parents were never diagnosed with cancer plus individuals whose parents were diagnosed when they were below 15 years old (Panel A), 15-18 (Panel B), 19-24 (Panel C), 25-29 (Panel D), 30-39 (Panel E), 40-49 (Panel F) and 50-59 years old (Panel G). From Panel A we conclude that, for individuals whose parents were diagnosed with lung cancer, we find no significant difference in the likelihood of smoking with respect to individuals whose parents did not receive any diagnosis. In contrast, individuals whose parents were diagnosed with a smoking-related cancer before they were 15 years old are more likely to smoke than people with undiagnosed parents. Receiving one's own diagnosis of lung cancer significantly reduces the likelihood of smoking, while receiving one's own diagnosis of any other smoking-related cancer results in no significant effect.Conclusions are similar if we consider instead individuals whose parents were diagnosed with cancer when they were aged 15-18 (see Panel B). For all these young people below 18, we expect the diagnosis effect to dominate over the behavior transmission effect, since offspring have been exposed to the parents' behavior for a short time. However this is not the case, when considering lung cancer, we do not find significant different behavior between diagnosed and undiagnosed people, while when considering smoking-related cancer, people whose parents were diagnosed are more likely to smoke, pointing to the fact that the parent's diagnosis is not associated with the smoking behavior, since we can hardly believe that the behavior transmission effect dominates with so few years of exposure.

Panel C considers the sample including both people receiving the parent's diagnosis between 19 and 24 years old and people whose parents were never diagnosed with cancer. In this case, the parent's diagnosis does not display any significant coefficient whatever the type of cancer type considered. The parent's diagnosis does not significantly alter the smoking behavior of the offspring with respect to the behavior of people whose parents were not diagnosed. This suggests that the behavior transmission effect dominates over the diagnosis effect. Again, the main determinant of the individual smoking behavior are the fact of receiving one's own diagnosis of lung cancer, which strongly decreases the probability of smoking. Conclusions are similar if we consider a sample including individuals with any parent diagnosed when they were aged 25-29 and individuals with undiagnosed parents (see Panel D).

The situation is modified when we consider a sample including individuals with any parent diagnosed when they were aged 30-39 (see Panel E). In this case, independently on the parent's type of cancer (lung cancer or any other smoking-related cancer), we find a negative effect of the parent diagnosis. This implies that when the individual was between 30 and 39 at the time of the parent's cancer diagnosis, she is less likely to be a smoker than individuals whose parents were never diagnosed with cancer. Surprisingly, over this age category the diagnosis effect dominates over the behavior transmission effect. Our initial hypothesis according to which the relative importance of the behavior transmission effect over the diagnosis effect should increase along the life cycle does not hold. Again, receiving one's own diagnosis of lung cancer significantly reduces the probability of being a smoker while receiving one's own diagnosis of a smoking-related cancer has no significant effect.

Panel F considers the sample including people without parents' diagnosis and people whose parents where diagnosed with cancer when the individual was between 40-49 years old. Results are in line with those in Panels A and B. Having a parent diagnosed with lung cancer significantly decreases the probability of smoking with respect to individuals whose parents were not diagnosed. In contrast, when the parent is diagnosed with a smoking-related cancer no significant effect on the smoking probability arises. Again receiving an own diagnosis of lung cancer significantly decreases the probability of smoking, while no significant impact is found if the own diagnosis concerns a smoking-related cancer.

The last sample we consider includes people whose parents were never diagnosed and people whose parents were diagnosed when the individual was aged 50-59 (see Panel G). In this case we find that people whose parents were diagnosed with lung cancer are more likely to be smokers than people whose parents never received the diagnosis. This may reveal that when the parents diagnosis arrives too late in the life of the individual, her behavior concerning smoking is going to be driven by her past habits (acquired during childhood). Therefore, when we compute the average probability (along life) for these people to be smokers it is unsurprising that it is significantly larger than that of people whose parents had no diagnosis. For smoking-related cancer there is no significant effect of the parents' diagnosis. Again, the own diagnosis of lung cancer significantly decreases the probability of smoking while no significant effect is found if the own diagnosis concerns any other smoking-related cancer.

4.3.2 Estimations results from a fixed effects model: how past behavior drives present choices

Table 12 analyzes the individual probability of smoking given her past smoking behavior. We analyze then the influence of past risky behaviors on current risky behaviors. We consider the whole population, including both, people whose parents were diagnosed with any kind of cancer and people whose parents were not diagnosed. Columns (1)-(2) consider the case where the parents are diagnosed with a lung cancer, while columns (3)-(4) focus on other smoking-related cancers.

Unsurprisingly, columns (1) and (3) reveal that individuals who used to be smokers in the past period are significantly more likely to be current smokers. Note though that the probability of smoking is significantly reduced for individuals whose parents have been diagnosed with lung cancer or with any other smoking-related cancer.

Results are slightly modified when considering the number of years of smoking instead of the past smoking dummy variable (see columns (2) and (4)). In line with previous findings, the more years of smoking, the higher the likelihood of current smoking. This higher probability of being a current smoker is though reduced if the individual's parents have been diagnosed with any type of cancer. This decrease does not manage to compensate the fact that individuals whose parents have been diagnosed with any kind of cancer are more likely to smoke.

Table 11: Average effect of parents' cancer diagnosis on smoking depending on the age of the diagnosis

	(1) Lung cancer	(2) Smoking-related cancer
Panel A: Diagnosis before 15		
Parents' diagnosis	0.047	0.036***
	(0.030)	(0.009)
Own diagnosis	-0.191***	0.005
	(0.036)	(0.005)
Observations	194708	172279
R-squared	0.633	0.633
Panel B: Diagnosis 15-18 years old		
Parents' diagnosis	-0.022	0.016*
	(0.025)	(0.009)
Own diagnosis	-0.191***	0.005
	(0.036)	(0.005)
Observations	194731	172164
R-squared	0.633	0.633
Panel C: Diagnosis 19-24 years old		
Parents' diagnosis	0.020	0.004
	(0.014)	(0.006)
Own diagnosis	-0.199***	0.004
	(0.036)	(0.005)
Observations	195125	174039
R-squared	0.633	0.633
Panel D: Diagnosis 25-29 years old		
Parents' diagnosis	-0.017	-0.004
	(0.013)	(0.005)
Own diagnosis	-0.191***	0.005
	(0.036)	(0.005)
Observations	195215	174372
R-squared	0.633	0.633
Panel E: Diagnosis 30-39 years old		
Parents' diagnosis	-0.017**	-0.015***
	(0.009)	(0.004)
Own diagnosis	-0.187***	0.001
	(0.036)	(0.005)
Observations	196492	179336
R-squared	0.633	0.633
Panel F: Diagnosis 40-49 years old		
Parents' diagnosis	-0.018*	-0.007
	(0.010)	(0.005)
Own diagnosis	-0.189***	0.005
	(0.036)	(0.005)
Observations	196030	177306
R-squared	0.633	0.633
Panel G: Diagnosis 50-59 years old		
Parents' diagnosis	0.048***	0.004
	(0.018)	(0.006)
Own diagnosis	-0.190***	0.004
	(0.036)	(0.005)
Observations	195163	174039
R-squared	0.633	0.633
Dependent variable	Smoker	Smoker
Year fixed effects	Yes	Yes
Individual fixed effects	Yes	Yes
Age controls	Yes	Yes

Source: Constances (2012-2019). Notes: Robust standard errors are in parentheses and clustered at the individual level. Statistical significance: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

Table 12: Average effect of parents' diagnosis of lung cancer on smoking - FE model

	(1)	(2)	(3)	(4)
	Lung cancer	Lung cancer	Smoking-related cancer	Smoking-related cancer
Smoker_{t-1}	0.863***		0.864***	
	(0.000)		(0.000)	
Parents' diagnosis	-0.000	0.076***	0.001**	0.075***
	(0.001)	(0.004)	(0.000)	(0.002)
$\text{Smoker}_{t-1} \times$ Parents' diagnosis	-0.006***		-0.008***	
	(0.002)		(0.001)	
Nb years smoking		0.008***		0.009***
		(0.000)		(0.000)
Nb years smoking $\times$ Parents' diagnosis		-0.007***		-0.007***
		(0.000)		(0.000)
Intercept	0.039***	0.507***	0.039***	0.506***
	(0.003)	(0.017)	(0.003)	(0.017)
Dependent variable	Smoker	Smoker	Smoker	Smoker
Year fixed effects	Yes	Yes	Yes	Yes
Individual fixed effects	Yes	Yes	Yes	Yes
4 age categories	Yes	No	Yes	No
8 age categories	No	Yes	No	Yes
Sample	All	All	All	All
Observations	200327	200327	200327	200327
R-squared	0.909	0.642	0.909	0.643

Source: Constances (2012-2019). Notes: Robust standard errors are in parentheses and clustered at the individual level. Statistical significance: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

5 Conclusion

In this paper, we analyze the impact of a parents' negative health shock on the offspring's risky behaviors.We analyze how having a parent diagnosed with lung cancer or smoking-related cancer affects the individual smoking behavior.

We argue here that it is crucial to consider the timing of the parents' diagnosis to fully investigate the issue. In line with the literature, we find no significant impact of the age of the parents' diagnosis on the smoking behavior. We do find that the probability to quit smoking significantly increases if one of the parents is diagnosed with a lung cancer when the offspring is between 15-18 years old. Finally, controlling for time invariant individual heterogeneity, we do not find significant differences in the probability to smoke between individuals whose parents have been diagnosed and individuals whose parents have not been diagnosed.

All in all, it seems difficult to provide any public health policy recommendation concerning parents behavior. The offspring risky behavior does not seem strongly affected by their parents negative health shock. Individuals essentially react to their own negative health shocks.

References

Adda, J., Bjorklund, A., and Holmlund, H. (2011). The Role of Mothers and Fathers in Providing Skills: Evidence from Parental Deaths. IZA Discussion Papers 5425, Institute of Labor Economics (IZA).

Alam, S. A. (2015). Parental health shocks, child labor and educational outcomes: Evidence from Tanzania. *Journal of Health Economics*, 44(C):161–175.

Arcidiacono, P., Sieg, H., and Sloan, F. (2007). Living Rationally under the Volcano? An Empirical Analysis of Heavy Drinking and Smoking. *International Economic Review*, 48(1):37–65.

Arora, K. (2016). How does dementia onset in parents influence unmarried adult children's wealth. *Social Science & Medicine*, 152(C):156–165.

Bünnings, C. (2017). Does new health information affect health behaviour? The effect of health events on smoking cessation. *Applied Economics*, 49(10):987–1000.

Cas, A., Frankenberg, E., Suriastini, W., and Thomas, D. (2014). The Impact of Parental Death on Child Well-being: Evidence From the Indian Ocean Tsunami. *Demography*, 51(2):437–457.

Case, A. and Ardington, C. (2006). The impact of parental death on school outcomes: Longitudinal evidence from South Africa. *Demography*, 43(3):401–420.

Chen, S. H., Chen, Y.-C., and Liu, J.-T. (2009). The Impact of Unexpected Maternal Death on Education: First Evidence from Three National Administrative Data Links. *American Economic Review*, 99(2):149–153.

Darden, M. (2017). Smoking, Expectations, and Health: A Dynamic Stochastic Model of Lifetime Smoking Behavior. *Journal of Political Economy*, 125(5):1465–1522.

Darden, M. and Gilleskie, D. (2016). The Effects of Parental Health Shocks on Adult Offspring Smoking Behavior and Self-Assessed Health. *Health Economics*, 25(8):939–954.

Decker, S. and Schmitz, H. (2016). Health shocks and risk aversion. *Journal of Health Economics*, 50:156–170.

Göhlmann, S., Schmidt, C. M., and Tauchmann, H. (2010). Smoking initiation in Germany: the role of intergenerational transmission. *Health Economics*, 19(2):227–242.

Hillebrandt, M.-A. (2022). Impact of changes in relationship status on smoking behavior and body weight. *Economics & Human Biology*, 44:101077.

Johnston, D. W., Schurer, S., and Shields, M. A. (2013). Exploring the intergenerational persistence of mental health: Evidence from three generations. *Journal of Health Economics*, 32(6):1077–1089.

Khwaja, A., Silverman, D., Sloan, F., and Wang, Y. (2009). Are mature smokers misinformed? *Journal of Health Economics*, 28(2):385–397.

Khwaja, A., Sloan, F., and Chung, S. (2006). Learning about individual risk and the decision to smoke. *International Journal of Industrial Organization*, 24(4):683–699.

Le, H. T. and Nguyen, H. T. (2017). Parental health and children's cognitive and noncognitive development: New evidence from the longitudinal survey of Australian children. *Health Economics*, 26(12):1767–1788.

Le, H. T. and Nguyen, H. T. (2018). The Impact of Maternal Mental Health Shocks on Child Health: Estimates from Fixed-Effects Instrumental Variables Models for Two Cohorts of Australian Children. *American Journal of Health Economics*, 4(2):185–225.

Li, C. and Gilleskie, D. B. (2021). The influence of endogenous behaviors among social pairs: Social interaction effects of smoking. *Journal of Health Economics*, 80:102547.

Loureiro, M. L., Sanz-de Galdeano, A., and Vuri, D. (2010). Smoking habits: like father, like son, like mother, like daughter? *Oxford Bulletin of Economics and Statistics*, 72(6):717–743.

Morin, A. J. S., Moullec, G., Maïano, C., Layet, L., Just, J.-L., and Ninot, G. (2011). Psychometric properties of the Center for Epidemiologic Studies Depression Scale (CES-D) in French clinical and nonclinical adults. *Rev Epidemiol Sante Publique*, 59(5):327–340.

Pan, J. and Han, W. (2017). Exploring the intergenerational persistence of health behaviour: an empirical study of smoking from China. *BMC Public Health*, 17(1):1–9.

Schmidt, C. M. and Tauchmann, H. (2011). Heterogeneity in the intergenerational transmission of alcohol consumption: a quantile regression approach. *Journal of Health Economics*, 30(1):33–42.

Senne, J.-N. (2014). Death and schooling decisions over the short and long run in rural Madagascar. *Journal of Population Economics*, 27(2):497–528.

Sloan, F. A., Smith, V. K., and Taylor, D. H. (2003). *The smoking puzzle: Information, risk perception, and choice*. Harvard University Press.

Smith, V. K., Taylor, D. H., Sloan, F. A., Johnson, F. R., and Desvousges, W. H. (2001). Do Smokers Respond To Health Shocks? *The Review of Economics and Statistics*, 83(4):675–687.

Sundmacher, L. (2012). The effect of health shocks on smoking and obesity. *The European Journal of Health Economics*, 13(4):451–460.

Zins, M., Goldberg, M., and CONSTANCES team (2015). The french constances population-based cohort: design, inclusion and follow-up. *European Journal of Epidemiology*, 30(12):1317–1328.

A Appendix: Descriptive statistics

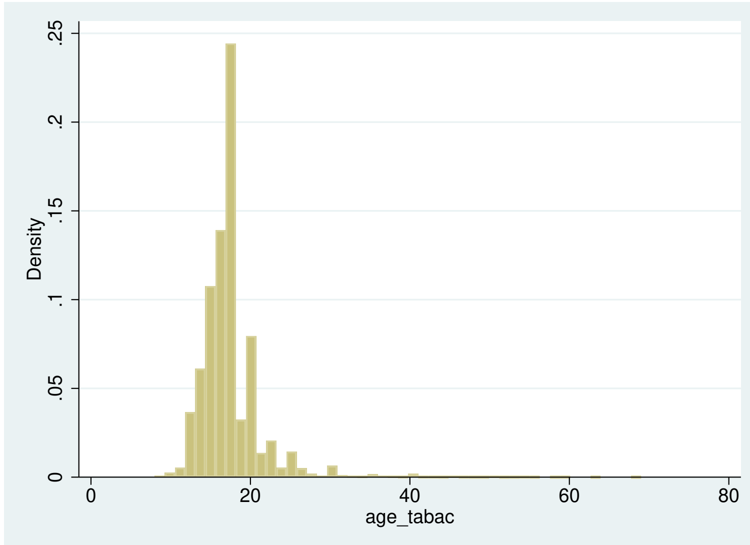
A.1 Sample composition

Table A.1: Sample composition 2012-2018. Socio-demographics

Variable Names	Obs	All	Obs	Male	Obs	Female
Female	92619	0.534				
15-29 years old	20335	0.117	8610	0.106	11725	0.127
30-39 years old	37581	0.216	17182	0.212	20399	0.220
40-49 years old	42597	0.245	19746	0.244	22851	0.247
50-59 years old	36459	0.210	17068	0.211	19391	0.209
60+ years old	36626	0.211	18373	0.227	18253	0.197
No education	4048	0.023	2194	0.027	1854	0.020
CFG/CEP/Brevet/BEPC	8434	0.049	3737	0.046	4697	0.051
CAP/BEP	25772	0.148	14780	0.183	10992	0.119
Bac	28780	0.166	12877	0.159	15903	0.172
Bac +2 or 3	46861	0.270	18699	0.231	28162	0.304
Bac +4	15066	0.087	5745	0.071	9321	0.101
Bac +5 and above	44207	0.255	22759	0.281	21448	0.232
Other	430	0.002	188	0.002	242	0.003
More than BAC	106564	0.614	47391	0.585	59173	0.639
In a relationship	128175	0.738	62195	0.768	65980	0.712
With children	83195	0.479	37490	0.463	45705	0.493
French	160180	0.928	74615	0.927	85565	0.929
French by naturalization	6946	0.040	3268	0.041	3678	0.040
Foreign	5460	0.032	2596	0.032	2864	0.031
Immigrant	17642	0.102	8491	0.105	9151	0.099
Second generation immigrant	6233	0.036	2920	0.036	3313	0.036
Metropolitan France	155956	0.898	72488	0.895	83468	0.901
Dom-Tom	1506	0.009	650	0.008	856	0.009
Europe	6968	0.040	3255	0.040	3713	0.040
North Africa	4269	0.025	2338	0.029	1931	0.021
Sub-saharian Africa	1726	0.010	945	0.012	781	0.008
Asia	1333	0.008	573	0.007	760	0.008
Other	1697	0.010	667	0.008	1030	0.011
NA	143	0.001	63	0.001	80	0.001
Native	155956	0.925	72488	0.923	83468	0.927
0-9 years	2343	0.014	1080	0.014	1263	0.014
10-19 years	3114	0.018	1419	0.018	1695	0.019
20-29 years	1931	0.011	898	0.011	1033	0.011
30+ years	5246	0.031	2674	0.034	2572	0.029
Employed	119537	0.689	56139	0.693	63398	0.685
Unemployed	10080	0.058	4482	0.055	5598	0.060
Retired	29922	0.172	15247	0.188	14675	0.158
No working activity	5726	0.033	2120	0.026	3606	0.039
Student, trainee	1577	0.009	738	0.009	839	0.009
Disability	3314	0.019	713	0.009	2601	0.028
Other	3442	0.020	1540	0.019	1902	0.021
High income	102094	0.588	50523	0.624	51571	0.557

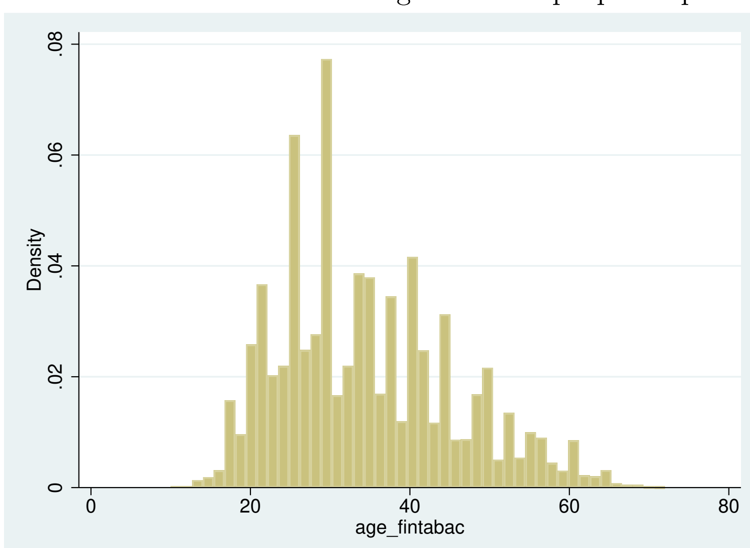
Source: Constances (2012-2018).

Figure A.1: Distribution of the age at which people start smoking



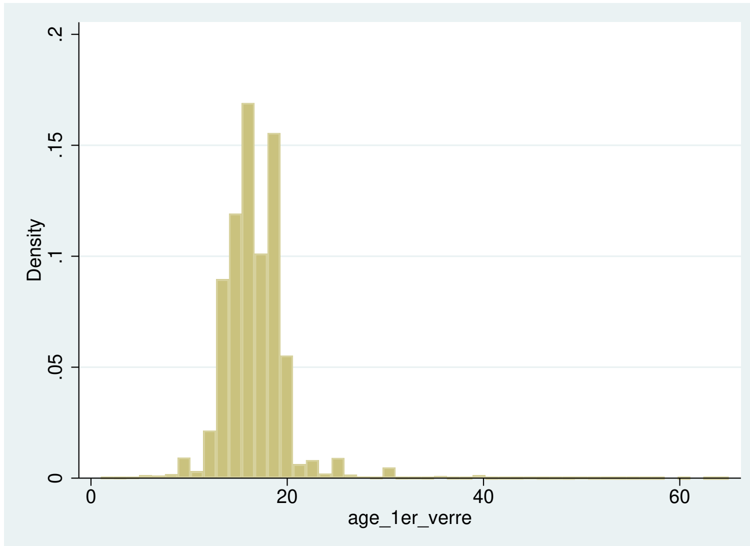
Source: Constances (2012-2018). Authors computations.

Figure A.2: Distribution of the age at which people stop smoking



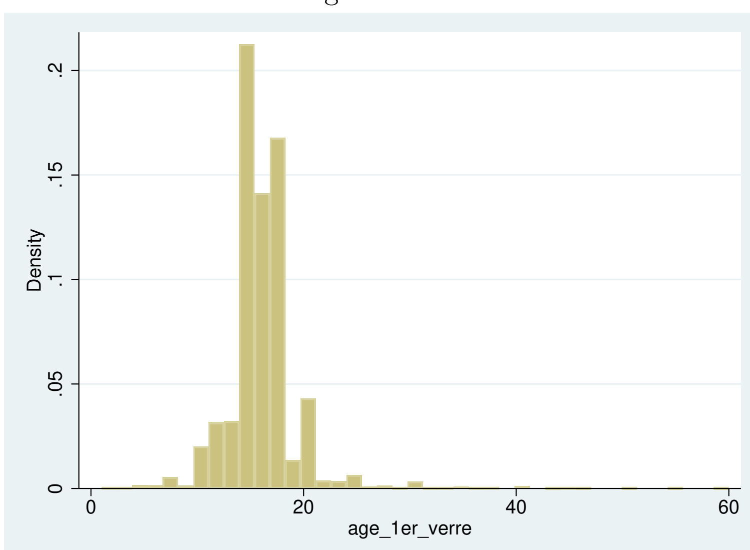
Source: Constances (2012-2018). Authors computations.

Figure A.3: Distribution of the age at which moderate drinkers started drinking



Source: Constances (2012-2018). Authors computations.

Figure A.4: Distribution of the age at which excess drinkers started drinking



Source: Constances (2012-2018). Authors computations.

Table A.2: Socio-demographics differences between not longer smokers and smokers at the first wave, 2012-2018.

Variable Names	Mean Non-smokers	Mean Smokers	P-value
Female	0.543	0.509	0.000
Age	48.107	42.379	0.000
15-29 years old	0.101	0.175	0.000
30-39 years old	0.192	0.274	0.000
40-49 years old	0.233	0.251	0.000
50-59 years old	0.215	0.189	0.000
60+ years old	0.259	0.112	0.000
No education	0.023	0.039	0.000
CFG/CEP/Brevet/BEPC	0.054	0.052	0.054
CAP/BEP	0.155	0.172	0.000
Bac	0.159	0.194	0.000
Bac +2 or 3	0.264	0.261	0.286
Bac +4	0.088	0.073	0.000
Bac +5 and above	0.255	0.207	0.000
Other	0.002	0.003	0.337
In a relationship	0.754	0.636	0.000
With children	0.478	0.477	0.749
French	0.925	0.933	0.000
French by naturalization	0.042	0.036	0.000
Foreign	0.033	0.032	0.373
Immigrant	0.103	0.103	0.698
Second generation immigrant	0.035	0.039	0.001
Employed	0.661	0.707	0.000
Unemployed	0.049	0.098	0.000
Retired	0.215	0.085	0.000
No working activity	0.030	0.045	0.000
Student, trainee	0.008	0.017	0.000
Disability	0.020	0.021	0.055
Other	0.018	0.028	0.000

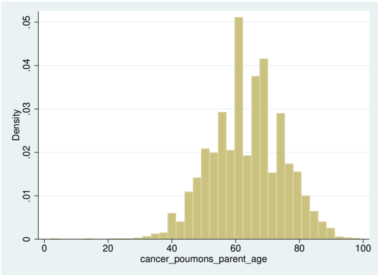
Source: Constances (2012-2018).

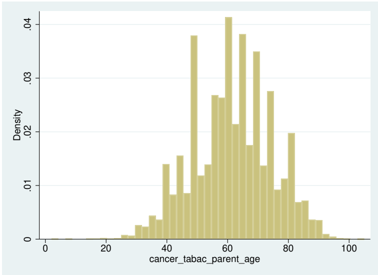
Table A.3: Socio-demographics differences between non excess alcohol drinkers and excess alcohol drinkers at the firs wave, 2012-2018.

Variable Names	Mean Non-excess drinkers	Mean Excess-drinkers	P-value
Female	0.539	0.412	0.000
Age	46.908	48.152	0.000
15-29 years old	0.110	0.130	0.000
30-39 years old	0.211	0.194	0.000
40-49 years old	0.244	0.183	0.000
50-59 years old	0.214	0.195	0.000
60+ years old	0.221	0.297	0.000
No education	0.021	0.026	0.000
CFG/CEP/Brevet/BEPC	0.049	0.057	0.000
CAP/BEP	0.150	0.165	0.000
Bac	0.163	0.161	0.452
Bac +2	0.270	0.243	0.000
Bac +4	0.087	0.085	0.278
Bac +5 and above	0.256	0.260	0.284
Other	0.002	0.003	0.161
In a relationship	0.745	0.691	0.000
With children	0.489	0.367	0.000
French	0.938	0.933	0.006
French by naturalization	0.035	0.032	0.066
Foreign	0.027	0.035	0.000
Immigrant	0.089	0.099	0.000
Second generation immigrant	0.033	0.028	0.000
Employed	0.692	0.584	0.000
Unemployed	0.052	0.076	0.000
Retired	0.182	0.251	0.000
No working activity	0.031	0.033	0.133
Student, trainee	0.008	0.012	0.000
Disability	0.018	0.018	0.887
Other	0.018	0.026	0.000

Source: Constances (2012-2018).

Figure A.5: Distribution of the age at which parents were diagnosed with smoke-related diseases





Source: Constances (2012-2018). Authors computations.

B Appendix: Results on smoking duration

Table B.1: Influence of the age at which the individual's parents were diagnosed with smoking-related cancer on the individual's smoking duration: all smokers

	(1)	(2)	(3)	(4)
Hazard ratios	b/se	b/se	b/se	b/se
Under 15 years old	1.097*	0.961	0.961	0.958
	(0.055)	(0.050)	(0.051)	(0.051)
15-18 years old	1.114**	1.004	1.029	1.030
	(0.055)	(0.052)	(0.055)	(0.055)
19-24 years old	1.095**	0.995	1.014	1.015
	(0.042)	(0.040)	(0.042)	(0.042)
25-29 years old	1.086**	1.010	1.027	1.028
	(0.039)	(0.038)	(0.040)	(0.040)
30-39 years old	1.027	0.978	0.984	0.985
	(0.030)	(0.030)	(0.032)	(0.032)
40-49 years old	1.015	1.004	1.016	1.017
	(0.031)	(0.031)	(0.033)	(0.033)
Smoking starting year FE	No	Yes	Yes	Yes
Smoking starting age FE	No	Yes	Yes	Yes
Department FE	No	Yes	Yes	Yes
Demographics	No	No	Yes	Yes
Educational attainment	No	No	Yes	Yes
Employment-related controls	No	No	No	No
Health-related controls	No	No	No	No
Work-related controls	No	No	No	No
Sample	Smokers	Smokers	Smokers	Smokers
Model	CoxPH	CoxPH	CoxPH	CoxNPH
Observations	17417	17417	16405	348907
R-squared				