Impact of acute health shocks on cigarette consumption
Antoine Marsaudon, Lise Rochaix

To cite this version:
Impact of acute health shocks on cigarette consumption: A combined DiD-matching strategy to address endogeneity issues in the French Gazel panel data∗

Antoine Marsaudon† Lise Rochaix‡

October 26, 2017

Abstract

This paper investigates the relationship between an acute health shock, namely the first onset of an accident requiring medical care, and cigarette consumption, using the French Gazel panel data. To identify the causal effect of such shocks, we use a difference-in-differences approach combined with a propensity score. Results suggest that there is a significant effect running from the shock to the number of cigarettes smoked with impact duration of eight years after the shock. Individuals subject to such a shock smoke, on average, 2 cigarettes less (per week) than those not exposed to such a shock. Further, the findings show heterogeneous effects among smokers: heavy smokers are more likely to reduce tobacco consumption than occasional smokers.

Keywords: health shock, panel data, France, lifestyles, propensity score matching

JEL: C23, I10, I12

∗We give special thank to Célia Lemaire (Humanis, EM Strasbourg), to Carole Treibich (Grenoble, UMR GAEL), to Jean-Pierre Messe (GAINS-TEPP, Le Mans), to Thomas Barney (ERUDITE, Créteil), to Alexandre Sauquet (LEMETA), and to Andrew Jones (York University). We also thank all participants of the 13th Doctorissimes (Paris 1), of the 16th Louis-André Gérard-Varet, and of the 2017 iHEA. We gratefully acknowledge Noémie Kiefer, Léa Toulemon, Laurie Rachet-Jacquet for their feedbacks on panel data and on matching strategies. The authors express their thanks to EDF-GDF, especially to the Service Général de Médecine de Contrôle, and to the “Caisse centrale d’action sociale du personnel des industries électrique et gazière”. We also wish to acknowledge the Cohortes team of the Unit UMS 011 Inserm - Versailles St-Quentin University responsible for the GAZEL data base management. The GAZEL Cohort Study was funded by EDF-GDF and INSERM, and received grants from the “Cohortes Santé TGIR Program”, Agence nationale de la recherche (ANR) and Agence francaise de sécurité sanitaire de l’environnement et du travail (AFSSET).

†Corresponding author. Hospinnomics (PSE – École d’Économie de Paris, Assistance Publique des Hôpitaux de Paris – AP-HP), Paris 1 Panthéon Sorbonne University. Address: 1 Parvis Notre-Dame, Bâtiment 1, 5e Étage 75004 Paris, France. Tel: +33 6 30 00 20 32. E-mail: antoine.marsaudon@psemail.eu.

‡Hospinnomics (PSE – École d’Économie de Paris, Assistance Publique des Hôpitaux de Paris – AP-HP), Paris 1 Panthéon Sorbonne University. E-mail: lise.rochaix@psemail.eu.
1 INTRODUCTION

By investigating the relationship between an acute health shock (i.e. the first\(^1\) onset of an accident requiring medical cares) and lifestyles (i.e. cigarette consumption), this paper contributes to a better understanding of smoking patterns. Drawing on behavioral economics, the analysis considers the health shock experience as the provision of new and credible information, which can be used to update personal health risk beliefs and which may subsequently affect individuals’ lifestyles.

Negative health shocks may both have beneficial (i.e. leading to reduce cigarette consumption) and detrimental (i.e. leading to increase cigarette consumption) effects on lifestyles. Regarding beneficial effects, two channels may be defined. First, there may be an increased social pressure to quit smoking through more frequent interactions with health care professionals and/or the urging of family members. Second, since health shocks are strongly associated with labor market inactivity and disability (Garcia Gomez and Lopez Nicola (2006); Garcia-Gomez (2011); Garcia-Gomez et al. (2013); Jones et al. (2016); Trevisan and Zantomio (2016)), resulting in lower individual and household incomes (Riphahn (1999); Garcia Gomez and Lopez Nicola (2006); Garcia-Gomez et al. (2013)), individuals may reduce or quit smoking because of new financial constraints. Detrimental effects may arise, due to the medical properties of nicotine that may become increasingly important as individuals have to cope with post-traumatic stress and/or fear due to reduced life expectancy (Op den Velde et al. (2002); Kelly et al. (2015)). Determining the direction and the magnitude of the effect of health shocks on cigarette consumption is thus an empirical issue.

Several studies have shown that health shocks can induce healthy changes among British adults (Clark and Etilé (2002)), on middle aged and retired Americans (Smith et al. (2001); Falba (2005); Khwaja et al. (2006); Keenan (2009)), or on ageing Germans (Sundmacher (2012)). Some studies also offer theoretical foundations for these behavioral changes (Grossman (1972); Becker and Murphy (1988); Clark and Etilé (2002)). Although studies differ with respect to the mechanism explored, they all predict a positive correlation between a decline in individuals’ health and decisions to adopt healthier lifestyles. In addition, (Clark and Etilé (2002))’s learning model assumes that individuals can learn over time from non-

---

\(^1\)We, therefore, focus only on individual having one health shock.
personal experiences (spouse or friends facing a shock\textsuperscript{2}).

Very little, however, is known in France, either on the impact of such shocks on both cigarette consumption levels, or on the duration of this effect. This paper proposes to bridge this gap by improving upon the existing literature in three ways. First, the measure of health shock, i.e. the first onset of an accident requiring medical care is comparatively less noisy than previously used measures, with less severe endogeneity issues. Facing an accident is more precise than a serious decline in self-reported health status (Garcia Gomez and Lopez Nicola (2006); Garcia-Gomez (2011); Sundmacher (2012)), a drop in the level of health satisfaction (Riphahn (1999)) or hospital admission (Garcia-Gomez et al. (2013)). These measures might reflect very different health situations (e.g. chronic or acute health shocks, physical or psychological deteriorations). Further, our measure of shock might present less severe endogeneity issues. Accidents may suffer less from reverse causality than the onset of heart attack (Smith et al. (2001); Sahm (2012)). This last measure could be partly the consequences of individuals’ lifestyles. Second, data allows us to assess the impact of such a shock up to eight years after its occurrence. Most studies encompass shorter timelines, with the exception of (Clark and Etilé (2002))’s study covering 7 years. Third, we provide empirical evidence on how people learn from personal experiences to change lifestyle choices. By doing so, this paper contributes to understand health shocks as a nudge used to modify health behaviors.

We use a very rich French panel data (\textit{Gazel}\textsuperscript{3}), which covers 20,000 individuals (15,000 men and 5,000 women) working for the French electricity board (EDF-GDF) over the 1989 to 2014 period, for the whole of France. Using this yearly panel data highlights both inter-individual differences and intra-individual dynamics and helps capture part of the complexity of decisions in this domain.

To identify the causal effect of the accident, a matching based on pre-accident covariates and pre-accident outcomes is performed. Specifically, we compute a propensity score for facing a shock one year before its occurrence using a Probit estimation which includes: demographic (age, gender, and marital status), and socioeconomic indicators (monthly household income, personal educational attainment, father’s socioeconomic status, professional status,

\textsuperscript{2}See also Bala and Goyal (1998); Jones (1994).
\textsuperscript{3}See more on: \url{http://www.gazel.inserm.fr/en/}, and on Goldberg et al. (2006).
marital status, alcohol consumption, place of residence and self-reported health), along with the pre-outcome variable (number of cigarettes smoked). We then associate a treated individual (i.e. facing a health shock) with a control individual (i.e. not facing a health shock) based on this propensity score. Additionally, we restrict the analysis to observations within the common support range, and individuals with other types of shocks are excluded from sample and thus not included in the control group. Further, to take into account both unobserved time-invariant individual effects (e.g. race, genetic factors, or innate ability), and unobserved individual-invariant time effects (e.g. cigarettes’ relative prices, or anti-smoking campaigns\textsuperscript{4}) we combine this matching with a difference-in-differences (DiD, hereafter) approach.

Results suggest that there is a significant effect running from the shock to the number of cigarettes smoked, with impact duration of eight years after the shock. Individuals subject to such a shock smoke on average 2 cigarettes less per week than those who do not face a shock. The findings are robust to a series of robustness checks.

The paper proceeds as follows. Section 2 describes the data. Section 3 presents our empirical strategy. Section 4 shows the results and reports the effectiveness of the identification strategy through the robustness check. The last section concludes and highlights avenues for future research.

2 DATA

The Gazel dataset is a yearly panel with approximately 20,000 individuals throughout all regions of France. It provides 25 waves of individual data on health status, lifestyles, socioeconomic and occupational factors collected via a standardized questionnaire. This questionnaire is sent to all participants every year by mail. The cohort was set up in January 1989, with an invitation to participate sent to all GDF-EDF male employees aged from 40 to 50, and to all female employees aged 35 to 50. Invitation letters only mention a participation in a long-term health study to improve medical research. Less than 5% of the global cohort

\textsuperscript{4}Smoking in France was first restricted on public transport by the Loi Veil launched in 1976. Further restrictions were established in 1991 due to the Loi Evin. This law contains a variety of measures against alcoholism and tobacco consumption. On February 2007 smoking is ban from public places, such as offices, schools or restaurants.
had died (861 men, and 155 women) by the end of 2005. Further, only 126 subjects (0.6%) dropped out during the first 17 years of follow-up (1989-2005).

To measure the health shock, we rely on the first onset of an accident requiring medical care and examine whether individuals stop, start or reduce their cigarette consumption thereafter. Specifically, we use the following question: “over the last twelve months, have you ever had accidents that led to medical care?” The shock measure adopted here is a dummy variable set to one if individuals face such a shock, and zero otherwise. The outcome variable is the number of cigarettes smoked per day and is a continuous variable ranging from 0 to 57. We also exploit a broad set of covariates, in particular age, gender, household income, father’s socioeconomic status, individual educational attainment, occupational status, marital status, self-reported health, alcohol consumption, and place of residence.

3 EMPIRICAL STRATEGY

3.1 Design

We aim at estimating the impact of health shocks on cigarette consumption. Yet the treated individuals may have some specific characteristics not shared with those of the control group. For instance, since our measure of shock focuses on accidents (both work and traffic accidents), gender is a dummy that values one for women and zero for men.

Income is an index ranging from one (the poorest) to 10 (the richest). More precisely: 1 stands for "earn less than 991 euros"; 2 for "earn more than 991 euros but less than 1144 euros"; 3 for "earn more than 1144 euros but less than 1372 euros"; 4 for "earn more than 1372 euros but less than 1601 euros"; 5 for "earn more than 1601 euros but less than 1982 euros"; 6 for "earn more than 1982 euros but less than 2592 euros"; 7 for "earn more than 2592 euros but less than 3811 euros"; 8 for "earn more than 3811 euros but less than 4574 euros"; 9 for "earn more than 4574 euros but less than 6098 euros"; 10 for "earn more than 6098 euros".

Father’s socio-economic status contains six measures. Specifically, 1 stands for farmers; 2 for craftsman; 3 for chief executive officer or for executive; 4 for intermediary profession; 5 for employee; 6 for worker.

Individual level of education is coded as follow: 1 for "lower than high school degree"; 2 for "higher than high school degree".

Occupation status equals to 1 if the individual is employed; 2 if the individual is in sick leave; 3 if the individual is retired; 4 if the individual is retired but still working.

Family status is coded as follow: 1 stands for being single; 2 for being in couple; 3 for being divorced or separated.

Individuals who identify them as is very good health are coded 1 and those in very bad health are coded 8. Answers could rank from 1 to 8.

Alcohol consumption is ranked from 0 to 4. More precisely, 1 for occasional drinker; 2 regular drinker; 3 for frequent drinker; 4 heavy drinker.

Place of residence is coded as follow: 1 for rural; 2 for city center; 3 for suburb; 4 for remote city.
dents), only those with a private mode of transportation will be concerned by the latter. Second, blue and white collars do not have the probability to face work accidents. Skilled blue collars are more likely to face a professional accident than white collars. Individuals in the treatment group could be more risk-taking (e.g. smoke more, eat fatter foods, or be less cautious drivers). Clearly, the probability for these individuals to face such shocks is not random. Taking into account the non-randomness of the occurrence of health shocks calls for a quasi-experimental design. Such will be the empirical strategy adopted here, with a balancing score matching approach as a first step, and a difference-in-differences (DiD, hereafter) approach as a second step. This is similar to other studies trying to disentangle the impact of shocks on labor outcomes (Garcia Gomez and Lopez Nicola (2006); Garcia-Gomez et al. (2013); Barnay et al. (2015); Trevisan and Zantomio (2016); Jones et al. (2016)).

The first step entails matching treated and non-treated individuals based on a single variable called a balancing score. A balancing score is a function of $X$, denoted by $f(X)$, that must satisfy the following balancing assumption (1):

$$ T \perp\!\!\!\perp X \mid f(X) \tag{1} $$

This means that, conditional on the balancing score, the set of observables $X$ is independent of assignment to treatment ($T$). For observations with the same balancing score, the distribution of observables is the same among the treatment and the control group. One possible balancing score is the propensity score (PS, hereafter) matching. The PS is the probability for an individual to participate in the treatment given his or her observed characteristics $X$.

The first identifying assumption that underlies a propensity score matching is the as-
assumption of unconfoundedness (or the assumption of selection on observables). It assumes that selection to treatment is based only on observable characteristics (i.e. all variables that influence both treatment assignment and outcomes are observed). To ensure non-violation of the unconfoundedness assumption, all variables that influence both treatment assignment and the outcome variables must be included in the PS. This assumption can be written as (2):

\[
\begin{align*}
Y(0) & \perp \perp T \mid P(X) \forall X \\
Y(1) & \perp \perp T \mid P(X) \forall X
\end{align*}
\]

(2)

In equation (2), \((Y(0))\) refers to the outcomes for control group and \((Y(1))\) for the treatment group; \((T)\) equals one if individual are treated, and zero otherwise.

A further requirement is the common support (or the overlap) condition. It rules out the possibility of perfect predictability of the treatment given \(X\) (3):

\[0 < P(D = 1 \mid X) < 1\]

(3)

This ensures that individuals with the same \(X\) values have a positive probability of being both treated and non-treated (Heckman et al. (1999)).

If both unconfoundedness and common support assumptions hold, the PS estimates the average treatment on the treated (ATT, hereafter) (4):

\[
\tau_{ATT}^{PSM} = \mathbb{E}_{P(X)}[D = 1 \{E[Y(1) \mid D = 1, P(X)] - E[Y(0) \mid D = 0, P(X)]\}]
\]

(4)

The PS matching consists in computing the average difference between the mean outcome of treated individuals characterized by a specific PS, and the mean outcome of control individuals characterized by a similar PS. The propensity score matching implies pairing each treated individual \(i\) with (one or more) comparable control individual(s) \(j\). Specifically, we compute the matched outcomes (i.e. \(Y_j^{PSM}\), hereafter) using the weighted outcomes of the nearest neighbors \(j\) of a treated individual \(i\) (5):

\[
Y_j^{PSM} = \sum w_{ij} Y_j
\]

(5)

Where \(w_{ij}\) is the weight of control individual \(j\), with \(w_{ij} = 1\), and \(Y_j\) stands for the
outcome of control individual $j$ before the matching\textsuperscript{16}. Thus, the ATT is given by (6):

$$ATT = \frac{1}{N} \sum [Y_i - Y_{jPSM}^i]$$  \hspace{1cm} (6)

Where $N$ is the number of treated individuals in the sample for which a matched control individual exists.

By combining this propensity score matching with a DiD approach, we remove both unobservable individual specific effects which are constant over time, and common time effects. A common time effect could be tobacco prices, for instance. In France, since the mid-nineties, tobacco prices have more than doubled\textsuperscript{17} and since price elasticities of younger and older individuals are comparatively higher (Chaloupka and Wechsler (1997); Chaloupka and Warner (2000)), this must be explicitly taken into account. The validity of this empirical strategy relies on the assumption that, in the absence of the shock, the treated and the control group would have had the same time trends, relative to their outcomes. This is formally written as: (7):

$$\alpha_{DiD} = \left[ E(P_{i,b} = 1 \mid T = 1) - E(P_{i,a} = 1 \mid T = 1) \right] - \left[ E(P_{i,b} = 1 \mid T = 0) - E(P_{i,a} = 1 \mid T = 0) \right]$$ \hspace{1cm} (7)

Where $b$ stands for before shocks, and $a$ for after shocks.

### 3.2 Implementation

The matching strategy requires that all variables influencing both treatment assignment and the outcome variable should be included in the PS. We first compute a propensity score with a Probit estimation (Wooldridge (2000); Imbens and Wooldridge (2009); Lechner et al. (2011)). The Probit estimation contains demographic variables (age, gender, and marital status) and socioeconomic variables (monthly household income, personal and father’s socioeconomic status, professional status), as well as self-reported health, alcohol consumption, place of residence, and the number of cigarettes smoked per day.

\textsuperscript{16}For example, if the three neighbors for individual $i$ weight are 0.8; 0.1; and 0.1, with $Y_j$ respectively equals to 19; 20; and 17 then, $Y_{jPSM}$ is 18.9 cigarette smoked ((0.8*19)+(0.1*20)+(0.1*17)).

\textsuperscript{17}See more on: \url{http://en.ofdt.fr/BDD/publications/docs/eftaalu5.pdf}. 

8
We then compute the PS one year before the occurrence of the health shock for the treated individuals, and for all years for the control group. We then perform an exact matching based on the year before the occurrence of shocks. This ensures that individuals have the same probability to face shocks for that year. The treated individual is then matched with a non-treated individual based on his or her PS the year before the shock. This controls, among other sources of heterogeneity, for the fact that individuals at the end of the period will have had more exposure to tobacco prevention campaigns than those experiencing the shock at the beginning of the period. After matching, individuals in both groups therefore have the same probability of facing a shock one year before its occurrence. We apply a 3-nearest neighbors matching procedure that selects, for each treated individual, the 3 closest controls (i.e. those that have the closest propensity score). The choice of the nearest neighbors is bounded to the common support range\(^{18}\) (and thus respects the common support assumption). We calibrate the maximum difference in the propensity score between matched and control subjects to be at 0.001\(^{19}\). This ensures that matched individuals have very similar propensity scores. Further, matching is performed with replacement\(^{20}\). Allowing matching with replacement involves a trade-off between bias and variance. It increases the average quality of the matching (i.e. treated individuals are more likely to be matched with a control individual with the same PS), thereby decreasing the bias. However, it also reduces the number of distinct controls used to construct the counterfactual outcomes, and therefore, increases the variance of the estimator (Smith and Todd (2005)).

3.3 Matching quality

To assess the matching validity, we display the following tables and figures. Figure 1 shows the distributions of the propensity score for treated (continuous line) and control (dashed line) individuals before (left-hand side), and after (right-hand side) the matching procedure. While some overlap in distributions is visible before matching, post-matching distributions exhibit a better fit. Table 1 further documents the efficiency of the matching strategy. Before

\(^{18}\)Treatment observations whose PS is higher than the maximum or less than the minimum PS of the controls are discarded.

\(^{19}\)This means, for example, that a treated individual with a propensity score of 0.6720 is matched with an individual in the control group with a propensity score of 0.6721 or 0.6719.

\(^{20}\)The same potential control individual could serve as a nearest neighbor-matched control for more than a single treated individual.
matching, substantial and statistically significant differences exist between the treatment and the control groups in the means of the variables. After matching, none of the differences in average characteristics are statistically significant at any conventional level. Table 1 also provides some descriptive statistics of our sample. Individuals facing a health shocks are more likely to be an employee, with low level of education, and live in big cities. Further, they also are more likely to be a male aged between 50 to 55 years old. Figure 1 and Figure 2 display the region of common support to ensure that the overlap between both groups is sufficient to make comparisons. The histograms display the propensity scores for treatment and control cases. Control and treated individuals span the full range of the propensity scores, which gives further support to our empirical strategy.

The matching strategy relies on the assumption that only observable data drives the probability to face health shocks. This is, however, unlikely to be the case, due to omitted variables (e.g.: individuals’ readiness to adopt preventive strategies, or their risk and time preferences\(^{21}\)) in equation (5). These variables may both influence individuals’ lifestyles and their probability to have such a shock. Following the literature, it appears that this omitted variable problem is partly alleviated by the fact that individual preferences are correlated with marital status (Schmidt (2008)), educational attainment (Wardle and Steptoe (2003); Jaroni et al. (2004); Dom et al. (2006); Sahm (2012)), and gender (Byrnes et al. (1999); Siegrist et al. (2002); Daruvala (2007); Jusot and Khlat (2013)), all of which are used in the matching step.

4 RESULTS

4.1 Main results

Table 2 shows the effect of a health shock on cigarette consumption using the difference-in-differences approach combined with a propensity score matching, as described in equation 7. We report results as follows. Column 1 gives the difference between the average number of cigarettes smoked one year after the shock and one year before the shock. Column 2 provides the same difference but 2 years after the shock. Likewise, columns 3 to 7 report

\(^{21}\)See Goto et al. (2009); Fieulaine and Martinez (2010); Hall et al. (2012); Jusot and Khlat (2013) for a deeper look at the relationship between individual preferences and lifestyles.
the effect of health shocks from 3 to 7 years after the shock. Individuals facing shocks reduce their cigarette consumption by comparison with individuals who do not face such shocks. More precisely, individuals facing such shocks reduce their consumption, on average, by 2 cigarettes per week and this reduction seems to be become more important as time goes by. This could indicate that as addiction recedes, individuals are increasingly able to further reduce their cigarette consumption. This corroborates epidemiological findings showing that reducing cigarette consumption reduces dependency (McNeill (2004); Benowitz and Henningfield (2013); Begh et al. (2015)).

While facing a health shock reduces the number of cigarette smoked on average, significant disparities may exist in the population and need to be further documented. To study heterogeneous effects, we compare occasional and heavy smokers, the former being defined as smoking at most 5 cigarettes per day and the latter at least 15 cigarettes per day. The results (Table 3 and 4) show that heavy smokers are more likely to reduce their cigarette consumption, compared to occasional smokers, for whom no significant effect was found.

Our results are in line with other international studies. In the US, in the UK, and in Denmark heavy smokers are more likely to remain abstinent after trying to stop smoking than occasional smokers (Burns (2000); Godtfredsen et al. (2001); Kotz et al. (2012)). This may be because heavy smokers receive more advice from their GP than occasional smokers (Kotz et al. (2013)). Further, our findings contribute to a growing literature on the stability of health preferences (Craig et al. (2014); Ami et al. (2017); Bunn et al. (2006); Masanja et al. (2012)). Health preferences here refer to the individuals’ readiness to adopt healthy behaviors. Smokers could therefore be seen as individuals with low health preferences. If health preferences were stable over time, then smokers would not reduce their cigarette consumption after shocks. We observe the opposite, which may suggest individuals are not endowed with constant health preferences. Several explications could be offered to better understand this finding. First, the cost of smoking becomes more important for heavy smokers once the shock occurs. This finding is consistent with (Orphanides and Zervos (1995))’s theoretical model suggesting that individuals learn from their own experience. Second, the impact of shocks could also contribute to modify time preferences (i.e. the degree to which one values the future more than the present) and/or risk preferences. Health shocks could lead to increase individuals’ preferences for the future, which is in line with
previous research on the impact of information campaigns on smoking patterns (Kenkel (1991); Chaloupka and Wechsler (1997); Clark and Etilé (2002)). Health shocks may also increase risk aversion, nudging individuals to be more cautious about their health.

4.2 Robustness checks

Two robustness checks were performed to assess the sensitivity of our results to the matching strategy, by testing alternative types of matching strategy.

First, we use a radius matching approach (Dehejia and Wahba (2002)), which includes not only the nearest neighbors within the caliper, but also all comparison members within the caliper. Using a 0.5 radius, an individual in the treatment group is thus matched with all individuals in the control group with a propensity score within a 0.5 difference. By using all comparison individuals available within the radius, it extends the number of units when good matches are available, thereby reducing the risk of bad matches. Table 5 shows robust and significant results.

Second, we implement a Kernel matching (Heckman et al. (1997); Heckman et al. (1998)). This is a non-parametric matching estimator using weighted averages of all individuals in the control group to construct the counterfactual outcomes. In this case, an individual in the treatment group is matched with all individuals in the control group and control individuals who are closer to the treated individual are given higher weights. One major advantage of this approach is to lower the variance as more information is used. Further, to avoid bad matches, we only keep individuals belonging to the common support. Table 6 shows results in line with the propensity score matching.

Overall, our results are robust to these two alternative matching strategies, which indicates that the reduction in cigarette consumption could, thus, reasonably be attributed to the health shock experienced.

A further robustness check is carried out, comparing individuals who experience shocks at different times over the 25 years span. In order to take into account this heterogeneous situation, we split the group and run separate regressions for two subgroups: those who experience the shock before 1991 and those who experience it after 2004 (in Table 7). We find that the difference in cigarette smoked is higher for the first group. The fact that results differ across the two time periods is an indication of the existence of trend, with an
increasing level of information on tobacco toxicity. These findings validate the matching approach adopted in this paper.

5 CONCLUSION

The paper offers informative evidence on how French workers from the national Gaz and Electricity board have changed their tobacco intake following an acute health shock. The findings suggest that there is a significant effect running from the shock to the number of cigarettes smoked with impact duration of eight years after the shock. Individuals subject to a shock smoke on average 2 cigarettes less (per week) than those who did not face such a shock. Further, heavy smokers are more likely to reduce tobacco consumption than occasional smokers. More generally, our results suggest both that health shocks nudge individuals to change their health behaviors and that individuals do not seem to display stable health preferences.

Our results, nonetheless, face external and internal validity limitations. External limitation is related to the fact that the sample is not representative of the French population. We only have information for individuals working in the public sector and aged 36 to 75; Males are over represented (73,26%) by comparison to their general population share (48,41%). There are more blue collars in the Gazel data, compared to the French population (31% versus 12,8%)\(^{22}\). Internal validity limitation is due to unobserved heterogeneity. Individual preferences (i.e. risk and time preferences) are not observed, and are likely to influence both the probability of being treated, and of smoking. If these preferences are stable over time, the first differences of our DiD alleviates this problem. Further, we are not able to document the severity level in the health shock variable, and neither can we distinguish between traffic or work accident.

The policy implications of our research suggests designing information campaigns that are as close as possible to individuals’ own experiences, mimicking the effects of health shocks or relying on peers’ experience sharing. A possible illustration could be the campaign in Uruguay: pictures of newborn defects on cigarette packs will induce future mothers to quit smoking (Harris et al. (2015)). Likewise, our results seem to emphasize the fact that the time

\(^{22}\)See more on: https://www.insee.fr/fr/information/2383410.
at which information is released matters (in our case, after the shock). Indeed individuals may be more sensitive to preventive information once they experienced a negative health shock. Yet, the present analysis cannot inform as to the individual’s motivation behind the decision to reduce or stop smoking. To do so, detailed information on individuals’ preferences would be needed here, such as that collected on a regular basis in the innovation panel of the UK Understanding Society panel. This would help identify the precise pathways that influence these complex decisions.
References


Kotz, D., Willemsen, M. C., Brown, J., and West, R. (2013). Light smokers are less likely to receive advice to quit from their gp than moderate-to-heavy smokers: a comparison of


Wooldridge, J. M. (2000). *The initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity*. University of Bristol, Department of Economics.
6 TABLES & FIGURES

Figure 1: Distribution of the propensity scores by groups before and after the matching.

Note: this figure shows the distribution of the propensity scores and its density among the treated and the control group. It displays that before the matching there are some differences between these two groups. After the matching, however, the two groups seem to be more comparable. The distribution of the propensity score does not run from 0 to 1 as it is the result of the propensity score with an exact matching on the year before the health shocks.
Figure 2: Distribution of the propensity score among control and treatment groups.

Note: this figure displays that most of the treated individuals have a matched control with propensity score close to their own (most are less than a 0.005 differences).
<table>
<thead>
<tr>
<th>Variables</th>
<th>Before matching</th>
<th>After matching</th>
<th>P-value</th>
<th>% bias</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treated</td>
<td>Matched</td>
<td>P-value</td>
<td>Treated</td>
<td>Matched</td>
</tr>
<tr>
<td>Father’s profession</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Farmer</td>
<td>0.107</td>
<td>0.105</td>
<td>0.877</td>
<td>0.107</td>
<td>0.094</td>
</tr>
<tr>
<td>Artisan</td>
<td>0.123</td>
<td>0.122</td>
<td>0.953</td>
<td>0.123</td>
<td>0.116</td>
</tr>
<tr>
<td>Executive or manager</td>
<td>0.216</td>
<td>0.223</td>
<td>0.594</td>
<td>0.216</td>
<td>0.240</td>
</tr>
<tr>
<td>Intermediate profession</td>
<td>0.061</td>
<td>0.059</td>
<td>0.679</td>
<td>0.062</td>
<td>0.072</td>
</tr>
<tr>
<td>Employee</td>
<td>0.297</td>
<td>0.321</td>
<td>0.136</td>
<td>0.297</td>
<td>0.302</td>
</tr>
<tr>
<td>Worker</td>
<td>0.110</td>
<td>0.090</td>
<td>0.045</td>
<td>0.110</td>
<td>0.091</td>
</tr>
<tr>
<td>Educational attainment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Higher degree</td>
<td>0.267</td>
<td>0.270</td>
<td>0.866</td>
<td>0.886</td>
<td>0.876</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Couple</td>
<td>0.886</td>
<td>0.901</td>
<td>0.149</td>
<td>0.886</td>
<td>0.876</td>
</tr>
<tr>
<td>Divorced or separated</td>
<td>0.076</td>
<td>0.071</td>
<td>0.613</td>
<td>0.076</td>
<td>0.084</td>
</tr>
<tr>
<td>Household income</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6,500 to less than 7,500</td>
<td>0.002</td>
<td>0.007</td>
<td>0.857</td>
<td>0.002</td>
<td>0.003</td>
</tr>
<tr>
<td>7,500 to less than 9,000</td>
<td>0.022</td>
<td>0.032</td>
<td>0.112</td>
<td>0.022</td>
<td>0.015</td>
</tr>
<tr>
<td>9,000 to less than 10,500</td>
<td>0.085</td>
<td>0.098</td>
<td>0.199</td>
<td>0.085</td>
<td>0.088</td>
</tr>
<tr>
<td>10,500 to less than 13,000</td>
<td>0.115</td>
<td>0.113</td>
<td>0.904</td>
<td>0.115</td>
<td>0.114</td>
</tr>
<tr>
<td>13,000 to less than 17,000</td>
<td>0.216</td>
<td>0.194</td>
<td>0.114</td>
<td>0.216</td>
<td>0.204</td>
</tr>
<tr>
<td>17,000 to less than 25,000</td>
<td>0.287</td>
<td>0.265</td>
<td>0.145</td>
<td>0.287</td>
<td>0.305</td>
</tr>
<tr>
<td>25,000 and more</td>
<td>0.271</td>
<td>0.288</td>
<td>0.249</td>
<td>0.271</td>
<td>0.272</td>
</tr>
<tr>
<td>Professional status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sick leave</td>
<td>0.009</td>
<td>0.006</td>
<td>0.299</td>
<td>0.009</td>
<td>0.007</td>
</tr>
<tr>
<td>Retired</td>
<td>0.349</td>
<td>0.267</td>
<td>0.000</td>
<td>0.349</td>
<td>0.342</td>
</tr>
<tr>
<td>Retired with professional activities</td>
<td>0.018</td>
<td>0.015</td>
<td>0.365</td>
<td>0.018</td>
<td>0.015</td>
</tr>
<tr>
<td>Cigarette smoked</td>
<td>13.564</td>
<td>14.012</td>
<td>0.258</td>
<td>13.564</td>
<td>14.067</td>
</tr>
<tr>
<td>Self-reported health</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very bad</td>
<td>0.254</td>
<td>0.273</td>
<td>0.212</td>
<td>0.254</td>
<td>0.261</td>
</tr>
<tr>
<td>bad</td>
<td>0.344</td>
<td>0.314</td>
<td>0.057</td>
<td>0.344</td>
<td>0.319</td>
</tr>
<tr>
<td>Average</td>
<td>0.191</td>
<td>0.203</td>
<td>0.389</td>
<td>0.191</td>
<td>0.192</td>
</tr>
<tr>
<td>Sufficient</td>
<td>0.103</td>
<td>0.106</td>
<td>0.814</td>
<td>0.103</td>
<td>0.114</td>
</tr>
<tr>
<td>good</td>
<td>0.060</td>
<td>0.042</td>
<td>0.010</td>
<td>0.060</td>
<td>0.050</td>
</tr>
<tr>
<td>Good</td>
<td>0.004</td>
<td>0.008</td>
<td>0.256</td>
<td>0.004</td>
<td>0.009</td>
</tr>
<tr>
<td>Excellent</td>
<td>0.003</td>
<td>0.003</td>
<td>0.987</td>
<td>0.003</td>
<td>0.005</td>
</tr>
<tr>
<td>Sex</td>
<td>0.246</td>
<td>0.257</td>
<td>0.461</td>
<td>0.246</td>
<td>0.224</td>
</tr>
<tr>
<td>Age</td>
<td>53.196</td>
<td>53.451</td>
<td>0.219</td>
<td>53.196</td>
<td>53.107</td>
</tr>
<tr>
<td>Place of residence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Big cities</td>
<td>0.324</td>
<td>0.312</td>
<td>0.426</td>
<td>0.324</td>
<td>0.342</td>
</tr>
<tr>
<td>Suburb</td>
<td>0.066</td>
<td>0.084</td>
<td>0.059</td>
<td>0.066</td>
<td>0.085</td>
</tr>
<tr>
<td>Isolated city</td>
<td>0.139</td>
<td>0.146</td>
<td>0.523</td>
<td>0.139</td>
<td>0.115</td>
</tr>
<tr>
<td>Alcohol consumption</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occasional drinkers</td>
<td>0.479</td>
<td>0.459</td>
<td>0.224</td>
<td>0.479</td>
<td>0.493</td>
</tr>
<tr>
<td>Frequent drinkers</td>
<td>0.218</td>
<td>0.237</td>
<td>0.189</td>
<td>0.218</td>
<td>0.235</td>
</tr>
<tr>
<td>Heavy drinkers</td>
<td>0.113</td>
<td>0.126</td>
<td>0.263</td>
<td>0.113</td>
<td>0.116</td>
</tr>
</tbody>
</table>

**Note:** This table reports the balancing between the two groups. There is no statistical differences between groups after matching. The standardised % bias is measured as the difference of the means in the treated and non-treated as a percentage of the square root of the average of the sample variances in the treated and controls groups.
Table 2: Difference-in-differences with propensity score matching

<table>
<thead>
<tr>
<th>Difference in cigarette smoked</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$DiD_{(t-1,t+1)}$</td>
<td>-0.155*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.083)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$DiD_{(t-1,t+2)}$</td>
<td>-0.224**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.092)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$DiD_{(t-1,t+3)}$</td>
<td>-0.296***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.102)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$DiD_{(t-1,t+4)}$</td>
<td>-0.278**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.115)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$DiD_{(t-1,t+5)}$</td>
<td>-0.336***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.114)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$DiD_{(t-1,t+6)}$</td>
<td>-0.330**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.135)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$DiD_{(t-1,t+7)}$</td>
<td>-0.436***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.151)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$DiD_{(t-1,t+8)}$</td>
<td>-0.290*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.156)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Observations 80,557 74,966 69,513 64,155 58,913 53,745 48,694 43,757

Note: this table shows the effect of health shocks on cigarette consumption using a difference-in-differences combined with a propensity score matching. Individuals facing health shocks are more likely to reduce smoking. This effect lasts at least 8 years after the occurrence of such shocks. Standard deviations in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. 
Table 3: Heterogeneous effect:
Difference-in-differences with propensity score matching for occasional smokers

<table>
<thead>
<tr>
<th>Difference in cigarette smoked</th>
<th>(DiD_{(t-1,t+1)})</th>
<th>0.085</th>
<th>(0.119)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(DiD_{(t-1,t+2)})</td>
<td>0.056</td>
<td>(0.122)</td>
<td></td>
</tr>
<tr>
<td>(DiD_{(t-1,t+3)})</td>
<td>0.022</td>
<td>(0.132)</td>
<td></td>
</tr>
<tr>
<td>(DiD_{(t-1,t+4)})</td>
<td>-0.013</td>
<td>(0.137)</td>
<td></td>
</tr>
<tr>
<td>(DiD_{(t-1,t+5)})</td>
<td>-0.059</td>
<td>(0.149)</td>
<td></td>
</tr>
<tr>
<td>(DiD_{(t-1,t+6)})</td>
<td>-0.155</td>
<td>(0.167)</td>
<td></td>
</tr>
<tr>
<td>(DiD_{(t-1,t+7)})</td>
<td>-0.123</td>
<td>(0.191)</td>
<td></td>
</tr>
<tr>
<td>(DiD_{(t-1,t+8)})</td>
<td>-0.080</td>
<td>(0.153)</td>
<td></td>
</tr>
</tbody>
</table>

Observations 24,218 22,475 20,781 19,119 17,499 15,914 14,353 12,835

Note: this table shows the effect of health shocks on cigarette consumption using a difference-in-differences combined with a propensity score matching. Occasional smokers do not reduce their cigarette consumption after shocks. Standard deviations in parentheses. * p < 0.05, ** p < 0.01, *** p < 0.001.
Table 4: Heterogeneous effect:
Difference-in-differences with propensity score matching for heavy smokers

<table>
<thead>
<tr>
<th>Difference in cigarette smoked</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$DiD_{(t-1,t+1)}$</td>
<td>-0.315***</td>
</tr>
<tr>
<td></td>
<td>(0.132)</td>
</tr>
<tr>
<td>$DiD_{(t-1,t+2)}$</td>
<td>-0.452***</td>
</tr>
<tr>
<td></td>
<td>(0.144)</td>
</tr>
<tr>
<td>$DiD_{(t-1,t+3)}$</td>
<td>-0.559***</td>
</tr>
<tr>
<td></td>
<td>(0.161)</td>
</tr>
<tr>
<td>$DiD_{(t-1,t+4)}$</td>
<td>-0.434***</td>
</tr>
<tr>
<td></td>
<td>(0.191)</td>
</tr>
<tr>
<td>$DiD_{(t-1,t+5)}$</td>
<td>-0.483**</td>
</tr>
<tr>
<td></td>
<td>(0.188)</td>
</tr>
<tr>
<td>$DiD_{(t-1,t+6)}$</td>
<td>-0.475**</td>
</tr>
<tr>
<td></td>
<td>(0.167)</td>
</tr>
<tr>
<td>$DiD_{(t-1,t+7)}$</td>
<td>-0.558**</td>
</tr>
<tr>
<td></td>
<td>(0.258)</td>
</tr>
<tr>
<td>$DiD_{(t-1,t+8)}$</td>
<td>-0.459*</td>
</tr>
<tr>
<td></td>
<td>(0.270)</td>
</tr>
</tbody>
</table>

Observations 39,153 36,515 33,934 31,404 28,926 26,465 24,074 21,721

Note: this table shows the effect of health shocks on cigarette consumption using a difference-in-differences combined with a propensity score matching. Heavy smokers reduce their cigarette consumption after shocks. Standard deviations in parentheses. * p < 0.05, ** p < 0.01, *** p < 0.001.
Table 5: Robustness checks:
Difference-in-differences with Radius matching

<table>
<thead>
<tr>
<th>Difference in cigarette smoked</th>
<th>$DiD_{(t-1,t+1)}$</th>
<th>0.188***</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(0.072)</td>
</tr>
<tr>
<td>$DiD_{(t-1,t+2)}$</td>
<td>-0.248***</td>
<td>(0.079)</td>
</tr>
<tr>
<td>$DiD_{(t-1,t+3)}$</td>
<td>-0.318***</td>
<td>(0.086)</td>
</tr>
<tr>
<td>$DiD_{(t-1,t+4)}$</td>
<td>-0.283***</td>
<td>(0.099)</td>
</tr>
<tr>
<td>$DiD_{(t-1,t+5)}$</td>
<td>-0.320***</td>
<td>(0.098)</td>
</tr>
<tr>
<td>$DiD_{(t-1,t+6)}$</td>
<td>-0.308***</td>
<td>(0.117)</td>
</tr>
<tr>
<td>$DiD_{(t-1,t+7)}$</td>
<td>-0.329**</td>
<td>(0.134)</td>
</tr>
<tr>
<td>$DiD_{(t-1,t+8)}$</td>
<td>-0.176</td>
<td>(0.133)</td>
</tr>
</tbody>
</table>

Observations: 80,575   74,982   69,529   64,169   58,925   53,757   48,707   43,768

Note: This table shows robust coefficients of the effect of health shocks on cigarette consumption. Individuals facing health shocks are more likely to reduce smoking. Standard deviations in parentheses. * p < 0.05, ** p < 0.01, *** p < 0.001.
Table 6: Robustness checks:
Difference-in-differences with Kernel matching

<table>
<thead>
<tr>
<th></th>
<th>Difference in cigarette smoked</th>
</tr>
</thead>
<tbody>
<tr>
<td>$DiD_{(t-1,t+1)}$</td>
<td>-0.178**</td>
</tr>
<tr>
<td></td>
<td>(0.072)</td>
</tr>
<tr>
<td>$DiD_{(t-1,t+2)}$</td>
<td>-0.237***</td>
</tr>
<tr>
<td></td>
<td>(0.079)</td>
</tr>
<tr>
<td>$DiD_{(t-1,t+3)}$</td>
<td>-0.307***</td>
</tr>
<tr>
<td></td>
<td>(0.087)</td>
</tr>
<tr>
<td>$DiD_{(t-1,t+4)}$</td>
<td>-0.269***</td>
</tr>
<tr>
<td></td>
<td>(0.098)</td>
</tr>
<tr>
<td>$DiD_{(t-1,t+5)}$</td>
<td>-0.306***</td>
</tr>
<tr>
<td></td>
<td>(0.098)</td>
</tr>
<tr>
<td>$DiD_{(t-1,t+6)}$</td>
<td>-0.291**</td>
</tr>
<tr>
<td></td>
<td>(0.117)</td>
</tr>
<tr>
<td>$DiD_{(t-1,t+7)}$</td>
<td>-0.316**</td>
</tr>
<tr>
<td></td>
<td>(0.134)</td>
</tr>
<tr>
<td>$DiD_{(t-1,t+8)}$</td>
<td>-0.166</td>
</tr>
<tr>
<td></td>
<td>(0.133)</td>
</tr>
</tbody>
</table>

Observations 80,575 74,982 69,529 64,169 58,925 53,757 48,707 43,768

Note: this table shows robust coefficients of the effect of health shocks on cigarette consumption. Individuals facing health shocks are more likely to reduce smoking. Standard deviations in parentheses. * p < 0.05, ** p < 0.01, *** p < 0.001.
Table 7: Heterogeneous effect of health shocks:
Differences between groups before 1991 and after 2004

<table>
<thead>
<tr>
<th></th>
<th>Difference in cigarette smoked</th>
<th>Difference in cigarette smoked</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before 1991</td>
<td>After 2004</td>
</tr>
<tr>
<td>$DiD_{(t-1,t+1)}$</td>
<td>-0.694*</td>
<td>-0.172*</td>
</tr>
<tr>
<td></td>
<td>(0.391)</td>
<td>(0.102)</td>
</tr>
<tr>
<td>Observations</td>
<td>5,497</td>
<td>43,237</td>
</tr>
</tbody>
</table>

Note: This table reports heterogeneous effect of cigarette consumption between groups depending on time horizon. Individuals living in the beginning of the 90’s are less likely to be exposed to long and continous preventive (and information) campaings than individuals living after 2004. Therefore, the later could benefit from those policies which help them to reduce their tobacco consumption. * p < 0.05, ** p < 0.01, *** p < 0.001.