Check for
updates

Is it Time to Quit? Smoking Persistence and Self-Rated Health

Donatella Lanari[1] · Marta Pasqualini[2] · Luca Pieroni[3]

Abstract

This study evaluates the impact of smoking on self-rated health using a British cohort born in 1970 that was followed through adult life. Records were taken for this dataset many times; individual self-rated health was first recorded in 1996 at age 26, and afterward at ages 30, 34, and 42. The smoking rate over time determined membership in the groups of current-smokers, never-smokers, and former-smokers. Estimates showed that the current-smokers group produced an increase in the probability of being in poor health with respect to never-smokers of about 10 percentage points in the long term. This result was also consistent when we used former-smokers as the control group, or other model specifications. The baseline estimates were not contradicted by robustness checks and policy implications of these results were discussed.

Keywords Ageing · Smoking · Self-rated health · Life-course · UK

Introduction

The use of tobacco is generally considered to be the principal cause of various ill effects, particularly vascular, neoplastic, and respiratory tract diseases. Among men, smoking has been shown to shorten mortality by 7 to 10 years (Rogers and Powell-Griner 1991). In addition, factors such as poorer nutrition, lower socioeconomic status, or social capital, which are linked with smoking behavior (Islam et al. 2017), may affect quality of life, increasing morbidity (Pasqualini et al. 2017).

✉ Marta Pasqualini
   marta.pasqualini@upf.edu

1   Department of Medicine, University of Perugia, Perugia, Italy

2   Department of Political Science, Universitat Pompeu Fabra, Ramon Trias Fargas, 25-27
    08005 Barcelona, Spain

3   Department of Political Science, University of Perugia, Perugia, Italy

The efforts to increase public knowledge about the detrimental effects of smoking on health and the use of various public interventions to reduce smoking have had substantial success in decreasing smoking rates in Europe and the U.S.; however, they have had less success in Asian countries, where the smoking rate among men is still high. Recent evidence has also suggested a decline in health-compromising behaviors (such as smoking) that is linked to the economic crises and the subsequent recovery period in Europe (Asgeirsdttir et al. 2016).

One general characteristic of smokers is that their persistence in smoking remains high, even in those who intend to quit (U.S. Surgeon General Report 2014). The motivations for this behavior that are proposed in the literature draw on different fields. For example, economics literature found that the (partial) knowledge about adverse effects of smoking on health does not help smokers to quit smoking because they shortsightedly disregard the future onset of disease (Viscusi 1990). Instead, epidemiological literature suggests that smokers have no intention of quitting because smoking helps relieve mental stress or may fear weight gain (McCaul et al. 2006; Pieroni and Salmasi 2016).

Self-rated health (SRH) is the indicator most widely used to assess health, according to the individual's perception, since it captures overall health status (Idler and Benyamini 1997), and it is typically obtained by asking individuals to describe their health status. SRH is a useful tool for identifying individuals and groups at risk for poor health and for monitoring health changes in populations (de Bruin et al. 1996; Pirani and Salvini 2012). Several studies have confirmed the validity and reliability of this indicator of a person's general health and well-being (Lundberg 1996; Martikainen et al. 1999), although less precise and culturally influenced with respect to objective health measures. Moreover, SRH correlates highly with mortality (Jylh 2009; Idler and Benyamini 1997) and it is a robust indicator in predicting the need for healthcare (Fylkesnes 1993).

As smoking has been found to be the most important risk factor for SRH (see McClave et al. 2009; U.S. Surgeon General Report 2014), the patterns of this relationship is involved in evaluating the effects of the persistence of smoking over time. Clearly, the estimates could be biased by a heterogeneous individual's health perception (Schneider et al. 2012), as well as selection bias of individual characteristics interacting with smoking behavior (Ronchetti and Terriau 2019). People who believe themselves to be healthy know about the dangers of smoking, and have stopped (i.e., the healthy-quitting effect). However, some healthy people simply believes they can handle the potential negative effects of smoking, continuing to smoke even when they know the risks, under-estimating smoking behavior on perceived bad health. Instead, people who are self-assessed to be less healthy get the message about the dangers of smoking and have stopped (i.e., disease-quitting effect), but there are also many sick people who have not stopped for various reasons (McCaul et al. 2006); this leads to over-estimation of smoking behavior on perceived poor health.

The present paper was designed to examine the relationship between smoking and SRH, using a British cohort born in 1970. Our contribution investigated whether smokers were more likely to state their health is good or bad than non-smokers, exploiting the SRH records of this cohort at ages 26, 30, 34, and 42 and applying a difference-in-difference (DID) approach to control for unobservable confounders in linear probability models (LPM) and ordered logistic models (OLM). We used former-

smokers as a control group, in addition to the natural control group of non-smokers, to account for other sources of bias related to reverse causality or omitted variables and estimate a causal relationship between smoking and SRH. With the exception of Yen et al. (2010), previous studies evaluated simple associations between smoking and SRH. In addition, our approach allowed us to investigate these effects differently over the course of life, a limitation which was evident in the literature that used national randomized surveys (Heikkinen et al. 2008), providing basic information about the timing of public health information.

Data

Data are drawn from the 1970 British Cohort Study (BCS70). This database is an ongoing national longitudinal study that involves more than 17,000 individuals born in Great Britain in a given week in April 1970. This survey was carried out every four years (the last wave in 2012), providing more than 15,000 multidisciplinary variables, and it contains detailed information at the micro level on health and related behaviors.

We selected waves referring to the years 1996, 2000, 2004, and 2012 because SRH was recorded for the first time in 1996 when the cohort reached age 26. Table 1 lists the number of individuals for each age; the corresponding number of dropouts between ages is expressed as survival and attrition rates. As expected, the attrition rate was 15% between the waves corresponding to the ages $26 - 30$ and increased to 39% at the age of 42. Despite the relatively severe attrition, the main questions were whether attrition significantly influenced the measurement of the health status indicator and whether differences across smoking characteristics (i.e., current-smokers, never-smokers, former-smokers) generated bias in the estimations. In the robustness section, we will respond to these questions.

Study Variables and some Stylized Facts

SRH is a good proxy of health obtained by asking the individuals to describe their health status according to a five-point scale: 1 = very good, 2 = good, 3 = fair, 4 = poor,

Table 1 Follow-ups of the 1970 British Cohort Study

	Age 26	Age 30	Age 34	Age 42
Longitudinal Sample (N. individuals)	9003	7638	6302	5485
Survival rate		84%	69%	60%
Drop-outs (N. individuals)		1365	2701	3518
Attrition rate		15%	30%	39%
Smoking habit	Age 26	Age 30	Age 34	Age 42
Current-smokers (CS)	33%	25%	22%	14%
Never-smokers (NS)	66%	66%	66%	66%
Ratio S/NS	50%	38%	33%	21%
Former-smokers (FS)		8%	12%	20%
Obs.	9003	7638	6302	5845

and 5 = very poor. We merged the last two categories into a poor category, since the latter contains a small number of observations. We also constructed an SRH dichotomous variable which coded the very good and good health statuses with 0 and fair, poor, and very poor with 1. Smoking is measured with yes/no item whether the respondent in each occasion smoked any kind of tobacco (yes = 1, no = 0). Cohort evidences of adult smoking behavior are listed in Table 1. The first column lists the proportion between current-smokers and never- smokers at age 26, recording a ratio of these variables of 0.5 (0.33/0.66); the magnitude of this ratio tends to decrease slightly to 0.38 (0.25 current-smokers and 0.66 never-smokers) at age 30, while it becomes smaller at age 42 (0.21; 0.14/0.66), driven from the effect of quitting rate (0.20). It is worth noting that the quitting rate is quite relevant before age 30.

To compare the health of smokers and non-smokers, we also needed to know the share of former-smokers over the lifetime course, which affected the outcome variables by selection. Here, we present some stylized facts of smokers persistence for the BCS70 cohort. We used the dichotomous specification of the SRH variable in the modalities poor health (PH) and good health for smokers (S = 1) and non-smokers (S = 0). In the observed year 1996 (age 26), PH was conditionally expressed in the following equation:

$$\frac{E(PH|S=0)_{1996}}{E(PH|S=1)_{1996}} = \frac{0.05}{0.11} = 0.45. \tag{1}$$

This means that the probability of being in poor health for non-smokers was less than half that of smokers. Clearly, the perception of health status evolved with individual age. The perception of being in poor health increased less in percentage for smokers by selection, such as the ratio increases, becoming:

$$\frac{E(PH|S=0)_{2012}}{E(PH|S=1)_{2012}} = \frac{0.30}{0.43} = 0.69. \tag{2}$$

We next examine which features of the BCS70 cohort may partially explain such a difference in the spontaneous dynamics, to obtain corrected estimates for the relationship between smoking and health status.

Covariates

In this subsection, we describe the set of time-varying and invariant covariates included in our analysis, while their descriptive statistics are listed in the Appendix (Tables 7 and 8). Time-invariant characteristics are: gender (men vs women); ethnic group (Withe British vs others); country of residence (England; Wales; Scotland) and education that was assessed according to at what age the respondent lefts full education. Time-varying characteristics have been included in the model by taking their value at the baseline year (1996) and by adding two dummies with regard to positive and negative change over time. Specifically, we controlled for: marital status (married vs single, separated/divorced, widowed); social class which has been assessed according to the National Statistics Socio-economic Classification (Higher professional and managerial

occupations; Lower managerial and professional occupations; Intermediate occupations; Small employers and own account workers; Lower supervisory and technical occupations; Semi-routine occupations; Routine occupations; Never worked and long-term unemployed) and having any children (yes vs no).

Finally, health conditions have been measured by taking into account any depressive symptoms (yes vs no), obesity (body mass index > 30), respiratory diseases such as asthma and chronic bronchitis (yes vs no), alcohol consumption (1 if the respondents self-reported more than once/twice a week and 0 otherwise) and the presence of any other chronic diseases such as cancer or diabetes (yes vs no).

The Conceptual Model

The descriptive statistics showed that the quitting ratio over the total number of smokers climbed from 1996 to 2012, partially explaining the evolution of smoking behavior. Thus, to model the causal effects of smoking on health, we also included quitting rate such that the ratio in Eq. (1) can be expressed as:

$$\frac{E(PH|S=0)}{E(PH|S=1)} - \frac{1}{1-E(Q|S=1)} * \frac{E(PH|S=0)}{E(PH|S=1,\ Q=0)} \qquad (3)$$

where PH is defined above, $Q = 1$ for smokers that quit and $Q = 0$ for others, $E(Q|S = 1)$ is the probability of quitting smoking, while $E(PH|S = 1,\ Q = 0)$ is the probability of being in poor health upon not having quit. The latter characteristic defines the persistence of smoking.

The first term on the right-hand side of the Eq. (3) reflects the *incapacitation effect* of quitters in contributing to poor health perception of smokers (i.e., those who stopped smoking cannot contribute to poor health of smokers), while for non-smokers $E(Q|S = 1) = 0$ and $E(PH|S = 0, Q = 1) = E(PH|S = 0)$. The second term of the Eq. (3) depends on the causal effect of smoking behavior on the propensity to perceive PH status, as well as selection bias. Framing the problem within the potential outcome approach (Rubin 1974), the difference in (poor) SRH conditional on not quitting may be rewritten as:

$$\frac{E(PH^1|S=0)}{E(PH^0|S=1,Q=0)} = \frac{E(PH^1|S=0)}{E(PH^0|S=0)} * \frac{E(PH^0|S=0)}{E(PH^0|S=1,Q=0)} \qquad (4)$$

where PH^1 and PH^0 are the potential outcomes conditional on being non-smokers or being smokers.

The first term on the right-hand side of the Eq. (4) is the real causal effect, while through the second term we identify the selection bias. This is the main threat to identifying the causal effect $E(PH^1|S=0)/E(PH^0|S=0)$, because we observe PH^0 only when $S = 0$ and in general smokers and non-smokers have different characteristics to begin with. In the same vein, we could write Eq. (4) in terms of potential outcomes observing PH^1 when $S = 1$ and non-smokers are a control group.

Based on descriptive statistics listed in the online Appendix (Tables 7 and 8), current smokers tend to be less educated, with a lower share of traditional families (less married

and children) and have lower obesity rates. Some of these differences certainly reflect a direct effect of smoking rates (the fact that smokers can stop smoking given the worsening of chronic diseases); however, other significance differences, which arise from predetermined observable characteristics (age gender, work), point to an important role for selection. Our quasi-experimental design allows us to account for selection, which help to identify the causal effect of smoking behavior.

Empirical Specification and Identification

We examine whether the smoking persistence of current-smokers produces significant health perception of the individual poor status. To do this, we define a benchmark model in which $SRH_{i,t}$ is the discrete outcome measure of individual i at time t and in which some fraction of the population smokes. Individuals are observed over a period (t=0) and (t=1) in which $D_{i,t} = 1$ defines if a smoker in t=0 persists in this choice in t=1 (persistent smoker). $D_{i,t} = 0$ indicates if the individual decides to quit smoking or never smoked; these represent our control groups.

In view of the longitudinal nature of our dataset, we adopt the DID approach to obtain consistent estimates of the effect of smoking on SRH. This strategy controls for individual unobservable time-invariant characteristics and defines different control groups to take into account the bias induced by reverse causality or time-varying unobservable characteristics. We have seen above that the unbalanced longitudinal sample comprises 5485 individuals according to their smoking habit over time. While the current-smokers were those who used tobacco consumers in 1996, and that did not change their habit over time until 2012, so that all were maintained in the estimation sample, never-smokers and former-smokers that repeatedly changed their non-smoking/smoking status over time were excluded. The balanced longitudinal sample of 3,532 individuals, which we use in the empirical section, was built by 486 current-smokers, 2,349 never-smokers and 697 former-smokers.The empirical specification of the DID model is expressed by the following equation:

$$SRH_{i,t} = \alpha + \eta D_i + \delta T_t + \beta(D_i {*} T_t) + \Sigma\pi X_{i,t} + \epsilon_{i,t} \tag{5}$$

where $SRH_{i,t}$ is a decreasing measure of self-rated health of individuals i in time t in two modalities (i.e., dichotomous classification) or in the four modalities (i.e., ordinal classification). The coefficient η associated with D_i captures any pre-existing difference among current-smokers and control groups; the coefficient δ associated with T_t is a proxy for unobserved variables which may affect SRHs of current-smokers, former-smokers or never- smokers. The coefficient π indicates the effect of both time-varying and time-invariant covariates $X_{i,t}$ which may affect directly the SRH of smokers and control groups.

The effect of smoking on SRH is captured by the interaction between $D_i * T_t$, and esti-mated by the coefficient β, which also represent the DID estimate of the average treatment effect (ATE) in two observed periods:

$$\beta^{DID} = E\left[SRH_{0,1}|X\right] - E\left[SRH_{0,0}X\right] - \\ E\left[SRH_{1,1}|X\right] - E\left[SRH_{1,0}|X\right]. \tag{6}$$

Using this framework, we differentially estimate the effects of smoking over a lifetime. In particular, we estimate the impact on the SRH in the short term, after four years (year 2000, age 30), medium term, after eight years (year 2004, age 34) and long term, after sixteen years (year 2012, age 42).

Our identification strategy assumes that SRH variations between 1996 and successive time-periods, in which SRH is recorded for current-smokers, are affected by the spontaneous dynamic (i.e., time-specific component), while former-smokers are affected by quit- ting smoking and by the spontaneous dynamic. However, quitting decisions are found related to the individual health status, because some disease correlated with smoking may induce people to stop smoking. Alternatively, the decision to quit smoking may be seen as part of a more general attitude aimed at improving health (McCaul et al. 2006). In this case, estimates of the relationship between smoking and SRH would be biased by reverse causality. We control the magnitude of this effect by including some disease-smoking related variables in the vector X (i.e., respiratory disease, depression and other chronic diseases). Instead, never-smokers are individuals who are not affected by reverse causality in the same way as former-smokers, because we assume that their SRH do not affect significantly their smoking decision. Comparative causal estimations will be used to evaluate the magnitude of the reverse causality bias.

As argued above, if we are willing to assume that deviations from the parallel paths of DID depend solely on differences in observable characteristics, conditioning on such differences should remove all biases. We follow Abadie (2005) and preliminary weight observations by the (inverse) propensity score of assignment, to adjust for differences between groups.

Results and Discussion

We estimate propensity score weighting exploiting all information available in our data set by a logit regression of an indicator variable for the smoking groups on the covariates described in Tables 7 and 8 of the Appendix. The kernel density in Fig. 1 shows that there is a tail of individuals in the never-smokers whose estimated propensity score is close to zero, meaning they are very different (in terms of observable characteristics) from current-smokers. Less important appears to be the difference when use former-smokers as a control group. In particular, the inverse propensity score weighting reduces the importance of observations at zero, while it increases the weight of observations in both groups that lie in the middle of the distribution of the estimated propensity score.

Table 2 lists estimates of LPM and robust standard errors using the sample of propensity score matching to balance the smoking groups. In the first three columns, former-smokers are included as the control group. Current-smokers have a higher significant poor health perception, estimated around to 10 percentage points in the long-term with respect of being in good health. This result is also consistent when we use never-smokers as the control group, since the point estimates are within the confidence intervals.

Finally, we provide estimates by using the OLM weighted by propensity score matching. To disentangle the contribution of current-smokers to SRH outcome, Table 3 lists the marginal effects of β on the modalities of SRH. By using the group of former-smokers comparatively, we find that the contribution to SRH of the different modalities
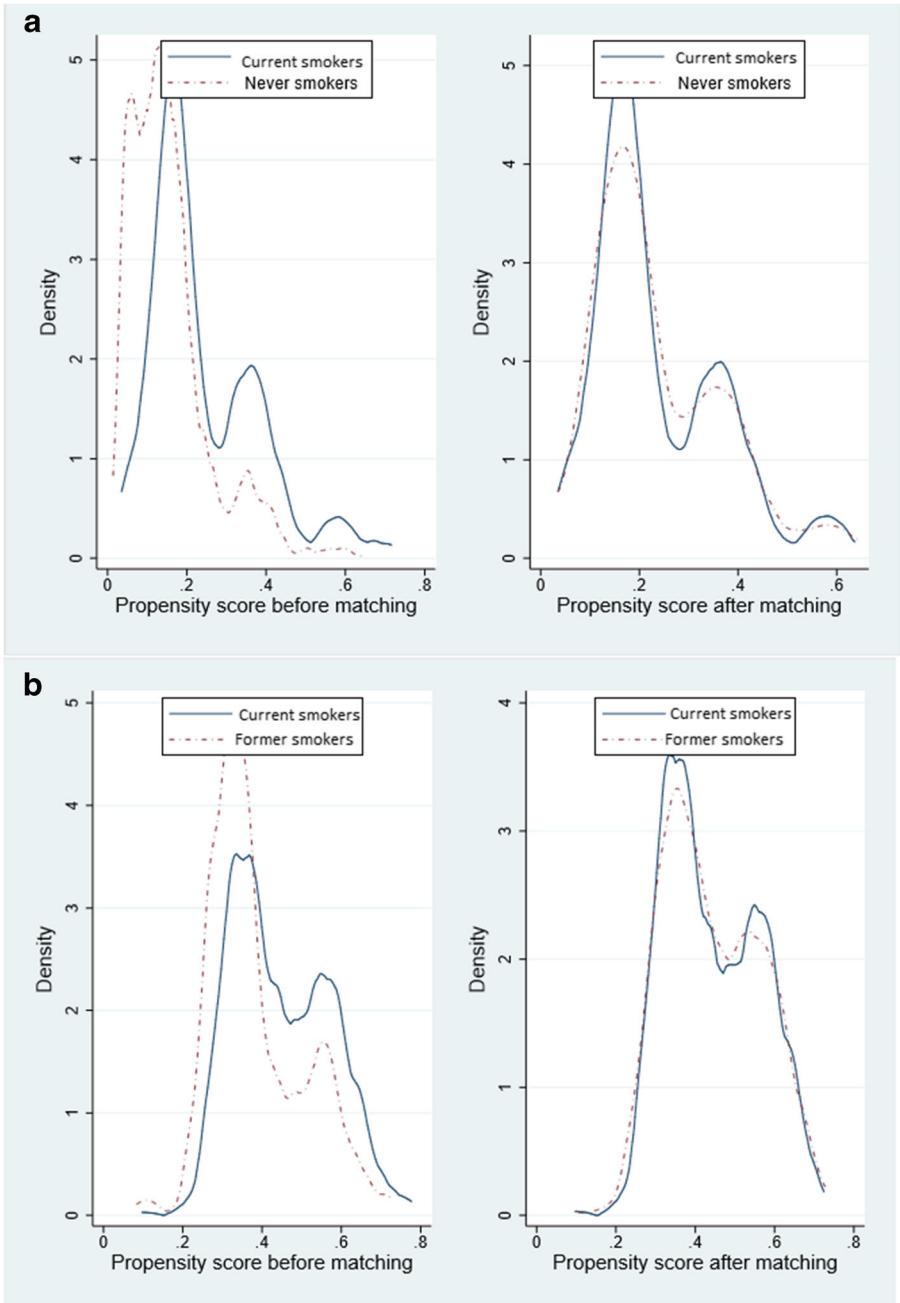
Fig. 1 Kernel density before and after the propensity-score matching. Note: Current smokers - never smokers (Panel **a**); Current smokers - Former smokers (Panel **b**)

is significant over time. Moreover, we find that the magnitude of the marginal effects using quitters is slightly large (and negative) in the "'very good" modality (i.e., 8.4

Table 2 Estimates of smoking persistence on self-rated health. Linear Probability Model

Coef.	Current smokers - Former smokers			Current smokers - Never smokers		
	Short-term	Medium-term	Long-term	Short-term	Medium-term	Long-term
η	0.121***	0.154***	0.212***	0.143***	0.177***	0.222***
	(0.0224)	(0.0192)	(0.0222)	(0.0139)	(0.0118)	(0.0134)
δ	0.0103	0.0241	0.0149	0.0417**	0.0576***	0.0503***
	(0.0274)	(0.0193)	(0.0157)	(0.0170)	(0.0118)	(0.00948)
β	0.0503**	0.0539**	0.104***	0.0376*	0.0278*	0.0959***
	(0.0247)	(0.0272)	(0.0313)	(0.0197)	(0.0166)	(0.0190)
Obs.	4732	4732	4732	11,340	11,340	11,340

The coefficients are related to the parameters of the Eq. 4. Estimates are controlled for time-invariant and time-varying covariates. Robust standard errors in parentheses. Significant levels as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Short-term = 2000; Medium-term = 2004; Long-term = 2012

percentage points; $SE = 0.032$) compared with the significant estimates of never-smokers in the long-term. Specifically, responses standardized by unitary variations of individual smokers that persist smoking over time are estimated to have a reduction of "very good" responses by 6.3 percentage points. The magnitude of these effects compared to the LPM estimates allows us to conclude that the choice of grouping the modality in a dichotomous way is highly consistent.

The Impact of Smoking Persistence by Individual Characteristics

Although our analyses mainly focus on the average effect of smoking persistence, it may be that it has little impact on most smokers but a very important impact on certain groups (Birch et al. 2005). Results in Table 4 for some significant covariates and illustrated for the long-term (1996-2012) reveal that smoking is sometimes different for any categorizations of the examined data. Corresponding findings for short and medium-term are remarkably in according with the trends we found in the baseline estimations.[1] While gender coefficient lists a poor health status for men although never statistically different at standard levels, we can reject the hypothesis that the effects are equal for the mutually exclusive groups concerning married people. In fact, the interval confidence of the analyzed category does not overlap completely. We also find that the effect is stronger for obese individuals. These findings suggest that continuing to smoke may thus be seen as part of a more general attitude towards bad health behavior.

Robustness and Parallel Trend Test

In this section, we examine the evidence that individuals that continue to smoke or quit smoking have different attitude about their actual and future health. The relevant questions are whether the strong habit of smoking, when begun prior to adulthood, may affect the

---

[1] The complete estimates for all covariates and for the short and medium period are available upon request from the authors.

Table 3 Marginal effects of smoking on self-rated health by control groups. Ordered Logistic Model

	Current smokers - Former smokers			Current smokers -Never smokers		
	Short-term	Medium-term	Long-term	Short-term	Medium-term	Long-term
Very Good	−0.080 ***	−0.073 ***	−0.084 ***	−0.006	0.006	−0.063 **
	(0.025)	(0.025)	(0.032)	(0.025)	(0.031)	(0.032)
Good	0.007 **	0.007 **	0.007 **	0.001	−0.001	0.016 **
	(0.003)	(0.003)	(0.004)	(0.006)	(0.006)	(0.008)
Fair	0.052 ***	0.047 ***	0.054 ***	0.003	−0.003	0.033 **
	(0.016)	(0.016)	(0.020)	(0.014)	(0.014)	(0.017)
Poor	0.020 ***	0.018 ***	0.022 ***	0.001	−0.001	0.013 **
	(0.006)	(0.006)	(0.008)	(0.005)	(0.005)	(0.007)
Obs.	4732	4732	4732	11,340	11,340	11,340

The table reports the β estimated coefficient for each modalities of self-rated health. Estimates are controlled for time-invariant and time-varying covariates. Robust standard errors in parentheses. Significant levels as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Short-term = 2000; Medium-term = 2004; Long-term = 2012

results we found, whether attrition may affect results and whether we do not reject the parallel trend assumption. Here, we present our results for the long-term SRH effects.

The first important question is whether starting to smoke at a young age affects the probability of continuing to smoke because of a reinforcement effect, which in turn has a negative effect on individual health. We extend the sample at age 16 where data whether individuals are smokers or non-smokers are available, while we do not know SRH at this age. In Table 5, we present the coefficient estimates for LPM, but the results are similar for OLM. As expected, the inclusion of the smoking habit variable increases the magnitude of the SRH coefficients in the long-term of 3–4 percentage points, irrespective of whether the control group is never-smokers or former-smokers, although we cannot reject that the confidence intervals of these estimates overlap those defined by the benchmark model.

The second important question concerns the influence of sample selection in the unbalanced longitudinal sample. We follow that of Verbeek and Nijman (1992) to compare results based on balanced and unbalanced panel data, modified to account for individual health status data which are missing because of non-response.[2] Indeed, if we assume that since people with poor health have a higher probability of dropping out during follow-up (Lanari et al. 2018), meaning that current-smokers report poorer health than never-smokers or former-smokers, the dropouts in this subgroup may lead to an underestimation of the decline in health status. We divided the longitudinal framework over the three periods which allows recovering many observations that may be missing over other periods between 1996 and 2012 and testing whether the sample design may have changed the point estimates when the missing value is replaced by the modality "poor health". Table 6 suggests that the estimates are close to those of the restricted final sample which includes individuals with observations that are present throughout the entire sample.

---

[2] Although we cannot know from our dataset why data are missing (nonresponse or death), the data from Global Burden Disease (1996–2017) suggest that death rates from smoking in the United Kingdom in the age 15–49 is on average under 0.05%.

Table 4 Estimates of smoking persistence on self-rated health by individual characteristics (Linear Probability Model)

Coef.	Current smokers - Former smokers			Current smokers - Never smokers		
	Short-term	Medium-term	Long-term	Short-term	Medium-term	Long-term
Gender						
Men	0.073*	0.116***	0.135***	0.041	0.048**	0.078***
	(0.041)	(0.036)	(0.041)	(0.028)	(0.023)	(0.026)
Women	0.005	−0.033	0.067	0.029	−0.001	0.100***
	(0.049)	(0.042)	(0.048)	(0.029)	(0.025)	(0.027)
Marital status: married						
Yes	0.052	0.077	0.178***	0.051	0.075***	0.222***
	(0.065)	(0.057)	(0.065)	(0.035)	(0.028)	(0.035)
No	0.049	0.042	0.072***	0.021	0.014	0.072***
	(0.036)	(0.031)	(0.023)	(0.024)	(0.020)	(0.023)
Children						
Yes	0.055	−0.021	0.113	−0.005	−0.024	−0.060
	(0.079)	(0.064)	(0.071)	(0.052)	(0.045)	(0.051)
No	0.072**	0.089***	0.127***	0.037*	0.046***	0.109***
	(0.034)	(0.029)	(0.034)	(0.021)	(0.017)	(0.021)
Obesity						
Yes	0.013	0.037	0.128*	0.058	0.04	0.103***
	(0.072)	(0.060)	(0.073)	(0.040)	(0.035)	(0.039)
No	0.057	0.073**	0.137***	0.017	0.034*	0.095***
	(0.035)	(0.030)	(0.036)	(0.022)	(0.019)	(0.022)
Alcohol use						
Yes	0.033	0.054*	0.118***	0.012	0.032	0.110***
	(0.035)	(0.030)	(0.035)	(0.023)	(0.019)	(0.022)
No	0.052	0.070	0.050	0.077*	0.037	0.061
	(0.070)	(0.059)	(0.066)	(0.041)	(0.034)	(0.038)
Obs.	4732	4732	4732	11,340	11,340	11,340

β coefficients were reported. Robust standard errors in parentheses. Time-varying variables were taken at the baseline (1996) and controlled for any positive and/or negative change over time. Significant levels as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Short-term = 2000; Medium-term = 2004; Long-term = 2012

The third important question investigates whether the causal effect of smoking on self-rated health is identified under the assumption that current-smokers and never-smokers or former-smokers have similar trends in terms of perceived health in the years immediately before the self-rated health registration in 1996. Unfortunately, we cannot test this assumption directly in our sample, because the 1996 was the first time in which data in SRH was collected. For this reason, we decided to complement this information relying on external sources of data. In particular, we obtained data on self-rated health by smoking behavior from the National Health Surveys (NHS) for the years 1992 and 1995 and selected the population for the cohort aged 22 and 25. These groups are approximately comparable and representative of young people in the UK. Figure 2 shows in

Table 5 Estimates of smoking persistence on self-rated health depending on early starting. Linear Probability Model (Long-term effects)

Coef.	Current - Former	Current - Never
η	0.202***	0.213***
	(0.026)	(0.014)
δ	0.027	0.054***
	(0.018)	(0.010)
β	0.139***	0.135***
	(0.037)	(0.020)
Obs.	3600	10,208

The coefficients are related to the parameters of the Eq. 4. Estimates are controlled for time-invariant and time-varying covariates. Robust standard errors in parentheses. Significant levels as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

both the trends of the self-rated health conditionally to the smoking habit using the NHS data for the period 1992-1995 (i.e., parallel trend assumption) and the BCS70 for the period 1996-2012. As we can see, from the graph there is definite evidence for the presence of a common trends in the period before the 26th birthday, irrespective if the control group are never-somokers or former-smokers, which supports our findings.

Limitations

The following factors deserve consideration and may also be seen as limitations of the paper. First, starting to smoke earlier represents a key element that determines smoking persistence over time and its effect on health. However, we were not able to evaluate this relationship by expanding the sample backward, because SRH is only evaluated starting from age 26. We can speculate that a large proportion of smokers began during adolescence (Amin and Lhila 2016). By exploiting information about smoking behavior at age 16, we recovered believable estimates for a sub-group of long-long-term smokers, suggesting significant and greater

Table 6 Selective attrition test of smoking persistence on self-rated health. Linear Probability Model (Long-term effect)

Coef.	Current - Former	Current -Never
η	0.241***	0.272***
	(0.0282)	(0.0169)
δ	−0.0114	0.0469***
	(0.028)	(0.0169)
β	0.130***	0.100***
	(0.0398)	(0.0239)
Obs.	2366	5670

The coefficients are related to the parameters of the Eq. 4. Estimates are controlled for time-invariant and time-varying covariates. Robust standard errors in parentheses. Significant levels as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

negative perceived health for this group of smokers, which is con- sistent with the epidemiological literature (see Strandberg et al. 2008). Moreover, this finding is particularly important in terms of policy implication since prior evidence indicates that young smokers have not the adequate level of access to smoking cessation services (Denscombe 2007) such as pharmacotherapy and behavioral support (Tonnesen 2009).

Second, our estimates use a smoking rate to obtain a binary dummy for smoking behavior. This measure may be not completely exhaustive of the effect on SRH because it does not account for the quantity of cigarettes smoked. Ho et al. (2003) showed that heavy smokers may have a better perception of health. In contrast, if consumption of more cigarettes, driven by nicotine addiction, is positively correlated with long-term smoking, predictions based on our estimates should suggest an even larger negative effect on perceived health.

Third, SRH is a global health measure that the literature has shown to depend on personal characteristics such as gender, education levels, or health behaviors. While the British cohort is sufficiently exhaustive concerning smoking behavior, correlated variables of health behavior, such as alcohol use or individual physical activity, do not have complete information. Clearly, it is difficult to estimate the potential bias. However, we are confident that the set of other control variables included in the analysis, which are generally correlated with health behavior, can minimize this issue.

Conclusion and Policy Recommendations

The present study exploited the long follow-up time (ages 30, 34, and 42) of the BCS70, which recorded SRH starting from age 26 (i.e., 1996), to show that current-smokers had a lower self-rated health; these differences were significant, irrespective whether we measured
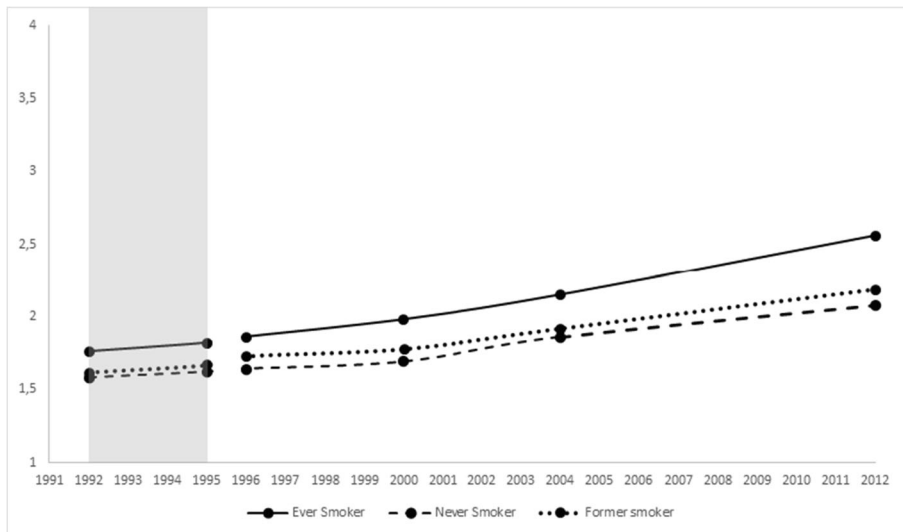

Fig. 2 Self-rated health by smoking characteristics over time. Source: Data for the pre-trend period (1992–1995) was drawn from the National Health Survey of England (NHS) available from IHME Burden of Disease

SRH in a dichotomous or an ordinal framework. Nevertheless, comparison over time of the SRH profiles of various smoking status groups pointed out where the differences lie: smokers perform worse on the SRH in the long-term compared with never-smokers, and the magnitude of this estimation increased at a decreasing rate.

Our study also examined the effect of smoking on SRH by including former-smokers as a control group. Clearly, the evaluation of the causal effect between smoking and SRH is complex because smoking is linked to many confounding factors, such as gender, educational level, and social status, while the reasons for cessation range from psychological factors to illness. The availability of data on several risk factors at the baseline allowed us to perform multivariate adjustment for potential confounding factors, including individual diseases.

These findings support the contention that long-term smoking cessation results in maintained or improved perceived health. Former-smokers tend to be more likely to report poor or fair health than never-smokers, particularly if they had recently quit smoking at the time of assessment (Ostbye et al. 2002; Arday et al. 2003). If we examine smoking patterns, the health consequences of smoking take some years to manifest themselves. Smokers showing early symptoms could have quit, leaving behind a more persistent group of smokers. The evidence that recent quitters had worse SRH is also consistent with reverse causation (i.e., that their deteriorating health was likely to be a major contributor to their quitting); the cohort of survivors in this study would miss the worst affected smokers, who might have died due to tobacco, or those with serious diseases or hospitalizations, who were excluded from the analysis.

In summary, beyond causing specific diseases and a wide range of other adverse health effects, smoking is also associated with generally poorer health; our findings support this.

Smoking adversely affects not only perceived health through specific disease pathogenesis and nonspecific mechanisms, such as alterations to the immune system, systemic oxidative stress, or subclinical organ injury, but also through personal psychological state, level of independence, and social relationships to essential features of the environment (Jylh 2009). The examination of the long term heterogeneous effects of smoking on SRH by individual characteristics showed that current-smokers did worse than never-smokers or former- smokers in all the dimensions. The findings of the present study were supported externally by those of Heikkinen et al. (2008) investigating the health-related quality of life profiles of healthy young smokers in Finland. They discovered that the never-smokers had higher mean of healthy-related quality of life scores than smokers in all the segments of demographic and socio-economic characteristics. When we took a more in-depth look at the significant mechanisms producing a heterogeneous effect of smoking on SRH, the British cohort did not show a statistically difference in coefficient magnitude of SRH between men and women. This result contrasts with that of Ho et al. (2003), based on a sample of individuals in Hong Kong Chinese, in which men smokers showed a greater and positive SRH. Significant effects have been found for variables directly linked with health, such as obesity or to be married, suggesting that their changes emphasized the point estimate differences of current-smokers with respect to never-smokers or former-smokers. According to Contoyannis and Jones (2004), efforts at education appear to produce health-positive results in the long-term.

Compliance with Ethical Standards   The authors declare that they have no competing interests and that no research funding were used to develop this article. The data for the British Cohort Study are managed by the Centre for Longitudinal Studies at the Institute of Education (University of London) and are available from the UK Data Service at the University of Essex. Individuals involved in the cohort have been guaranteed anonymity and all data collection on BCS70 has received full ethical approval (London Central REC). All authors read and approved the final manuscript.

Appendix

Table 7 Descriptive statistics - Shares of selected time-invariant covariates by groups

Variables	Current smokers	Never smokers	Former smokers	Total
Gender				
women	0.08	0.39	0.12	0.58
men	0.05	0.27	0.08	0.42
Ethnic group				
British	0.13	0.64	0.19	0.97
Other countries	0.005	0.02	0.005	0.03
Country				
England	0.12	0.56	0.17	0.86
Wales	0.01	0.04	0.01	0.06
Scotland	0.01	0.06	0.02	0.09
Education	0.16	0.18	0.17	0.17
Obs.	1944	7047	2788	14,128

Sources: The 1970 British Cohort Study. Time-invariant variables have been measured at the baseline year (1996). Note: Education is a dummy variable accounting for whether respondents left full education before (or at) age of 18

Table 8 Descriptive statistics - Shares of selected time-variant covariates by groups

Variables	Current smokers	Never smokers	Former smokers	Total
Marital status: married				
1996	0.03	0.23	0.04	0.31
2000	0.08	0.47	0.13	0.69
2004	0.06	0.42	0.10	0.60
2012	0.08	0.49	0.14	0.72
Social class: high level				
1996	0.14	0.17	0.17	0.17
2000	0.20	0.22	0.25	0.22
2004	0.08	0.22	0.20	0.20
2012	0.08	0.20	0.14	0.17
Obesity rate (BMI > 30)				

Table 8 (continued)

Variables	Current smokers	Never smokers	Former smokers	Total
1996	0.03	0.17	0.05	0.25
2000	0.04	0.21	0.09	0.34
2004	0.03	0.13	0.04	0.18
2012	0.04	0.21	0.06	0.31
Children				
1996	0.05	0.12	0.04	0.21
2000	0.06	0.27	0.08	0.41
2004	0.08	0.40	0.12	0.60
2012	0.08	0.45	0.14	0.67
Any respiratory disease				
1996	0.00	0.01	0.01	0.02
2000	0.01	0.01	0.05	0.07
2004	0.02	0.02	0.04	0.08
2012	0.01	0.01	0.05	0.08
Diabetes and cancer				
1996	0.00	0.00	0.00	0.00
2000	0.00	0.00	0.00	0.01
2004	0.00	0.01	0.00	0.01
2012	0.00	0.01	0.00	0.02
Depression				
1996	0.03	0.08	0.04	0.15
2000	0.03	0.08	0.06	0.17
2004	0.07	0.08	0.08	0.21
2012	0.04	0.11	0.04	0.19
Alcohol consumption				
1996	0.61	0.64	0.73	0.66
2000	0.59	0.70	0.72	0.70
2004	0.62	0.70	0.74	0.70
2012	0.51	0.54	0.41	0.51
	3			
Obs.	1944	7047	2788	14,128

Sources: The 1970 British Cohort Study. Social class is measured according to the UK Office for national Statistics's classification consisting of a eight classes variable (high level = Higher professional and managerial occupations)

References

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies, 72*(1), 1–19.

Amin, V., & Lhila, A. (2016). Decomposing racial differences in adolescent smoking in the US. *Economics & Human Biology, 22*, 161–176.

Arday, D. R., Milton, M. H., Husten, C. G., Haffer, S. C., Wheeless, S. C., Jones, S. M., & Johnson, R. E. (2003). Smoking and functional status among Medicare managed care enrollees. *American Journal of Preventive Medicine, 24*(3), 234–241.

Asgeirsdttir, T. L., Corman, H., Noonan, K., & Reichman, N. E. (2016). Lifecycle effects of a recession on health behaviors: boom, bust, and recovery in Iceland. *Economics & Human Biology, 20*, 90–107.

Birch, S., Jerrett, M., Wilson, K., Law, M., Elliott, S., & Eyles, J. (2005). Heterogeneities in the production of health: smoking, health status and place. *Health Policy, 72*(3), 301–310.

Contoyannis, P., & Jones, A. M. (2004). Socio-economic status, health and lifestyle. *Journal of Health Economics, 23*(5), 965–995.

de Bruin, A., Picavet, H. S. J., & Nossikov, A. (1996). *Health interview surveys. Towards international harmonization of methods and instruments*. Geneva: WHO, regional publications European Series no 58.

Denscombe, M. (2007). UK health policy and 'underage' smokers: the case for smoking cessation services. *Health Policy, 80*(1), 69–76.

Fylkesnes, K. (1993). Determinants of health care utilization–visits and referrals. *Scandinavian Journal of Social Medicine, 21*(1), 40–50.

Heikkinen, H., Jallinoja, P., Saarni, S. I., & Patja, K. (2008). The impact of smoking on health-related and overall quality of life: a general population survey in Finland. *Nicotine & Tobacco Research, 10*(7), 1199–1207.

Ho, S. Y., Lam, T. H., Fielding, R., & Janus, E. D. (2003). Smoking and perceived health in Hong Kong Chinese. *Social Science & Medicine, 57*(9), 1761–1770.

Idler, E. L., & Benyamini, Y. (1997). Self-rated health and mortality: a review of twenty-seven community studies. *Journal of Health and Social Behaviour, 38*(1), 21–37.

IHME Burden of Disease, http://ghdx.healthdata.org/gbd-results-tool

Islam, M. K., Folland, S., & Kaarbøe, O. M. (2017). Social capital and cigarette smoking: new empirics featuring the Norwegian HUNT data. *Economics & Human Biology, 26*, 174–185.

Jylh, M. (2009). What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Social Science & Medicine, 69*(3), 307–316.

Lanari, D., Bussini, O., & Minelli, L. (2018). The effects of immigrant status and age at migration on changes in older Europeans' health. *International Migration Review, 52*(4), 1218–1249.

Lundberg, U. (1996). Influence of paid and unpaid work on psychophysiological stress responses of men and women. *Journal of Occupational Health Psychology, 1*(2), 117–130.

Martikainen, P., Aromaa, A., Heliovaara, M., et al. (1999). Reliability of perceived health by sex and age. *Social Science and Medicine, 48*, 1117–1122.

McCaul, K. D., Hockemeyer, J. R., Johnson, R. J., Zetocha, K., Quinlan, K., & Glasgow, R. E. (2006). Motivation to quit using cigarettes: a review. *Addictive Behaviors, 31*(1), 42–56.

McClave, A. K., Dube, S. R., Strine, T. W., & Mokdad, A. H. (2009). Associations between health-related quality of life and smoking status among a large sample of US adults. *Preventive Medicine, 48*(2), 173–179.

Ostbye, T., Taylor, D. H., & Jung, S. H. (2002). A longitudinal study of the effects of tobacco smoking and other Modi?Able risk factors on ill health in middle-aged and old Americans: results from the health and

retirement study and asset and health dynamics among the oldest old survey. *Preventive Medicine, 34*(3), 334–345.

Pasqualini, M., Lanari, D., Minelli, L., Pieroni, L., & Salmasi, L. (2017). Health and income inequalities in Europe: what is the role of circumstances? *Economics & Human Biology, 26*, 164–173.

Pieroni, L., & Salmasi, L. (2016). The effect of smoking habit changes on body weight: evidence from the UK. *Economics & Human Biology, 20*, 113.

Pirani, E., & Salvini, S. (2012). Place of living and health inequality: a study for elderly Italians. *Statistical Methods & Applications, 21*(2), 211–226.

Rogers, R. G., & Powell-Griner, E. (1991). Life expectancies of cigarette smokers and nonsmokers in the United States. *Social Science & Medicine, 32*(10), 1151–1159.

Ronchetti, J., & Terriau, A. (2019). Impact of unemployment on self-perceived health. *The European Journal of Health Economics*, 1–11.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66*(5), 688–701.

Schneider, U., Pfarr, C., Schneider, B. S., & Ulrich, V. (2012). I feel good! Gender differences and reporting heterogeneity in self-assessed health. *The European Journal of Health Economics, 13*(3), 251–265.

Strandberg, A. Y., Strandberg, T. E., Pitkl, K., Salomaa, V. V., Tilvis, R. S., & Miettinen, T. A. (2008). The effect of smoking in midlife on health-related quality of life in old age: a 26-year prospective study. *Archives of Internal Medicine, 168*(18), 1968–1974.

Tonnesen, P. (2009). Smoking cessation: how compelling is the evidence? A review. *Health Policy, 91*, S15–S25.

U.S. Surgeon General Report. (2014). *The health consequences of smoking: 50 years of Progress*. Rockville: Department of Health and Human Services. Office of the Surgeon General.

Verbeek, M., & Nijman, T. (1992). Testing for selectivity bias in panel data models. *International Economic Review, 33*(3), 681–703.

Viscusi, W. K. (1990). Do smokers underestimate risks? *Journal of Political Economy, 98*(6), 1253–1269.

Yen, S. T., Shaw, W. D., & Yuan, Y. (2010). Cigarette smoking and self-reported health in China. *China Economic Review, 21*(4), 532–543.