Journal of Risk and Uncertainty (2022) 64:251–285
https://doi.org/10.1007/s11166-022-09378-9

Smoking, selection, and medical care expenditures

Michael E. Darden[1] · **Robert Kaestner**[2]

Accepted: 29 March 2022 / Published online: 30 April 2022

Abstract
The contribution of cigarette smoking to national health expenditures is thought to be large, but our current understanding of the effect of smoking on annual medical expenditures is limited to studies that use cross-sectional data to make comparisons of medical care expenditures between smokers and never smokers at a particular age. We develop a dynamic economic model of smoking and medical care use that highlights two forms of selection: selective mortality and non-random cessation. To test predictions from our model, we construct novel longitudinal profiles of medical expenditures of smokers and never smokers from merged National Health Interview Survey and Medicare claims information. Consistent with our theory, we find that, from a given age, smokers generate higher expenditures prospectively, because of a higher incidence in inpatient usage, and lower expenditures retrospectively, because of lower outpatient usage. Between ages 65 and 84, we find that the expected value of the discounted sum of total expenditures is *lower* for smokers, mainly because of excess mortality. We find no evidence that cigarette smoking is a burden on Medicare.

Keywords Health care expenditures · Medicare · Smoking

JEL I1 · I12 · I18

1 Introduction

Cigarette smoking remains the leading cause of preventable mortality in the United States (HHS, 2014). In 2017, there were roughly 34.2 million current smokers and 55.2 million former smokers in the United States over the age of 18 (Creamer et al.,

✉ Michael E. Darden
michaeldarden@jhu.edu

✉ Robert Kaestner
kaestner@chicago.edu

1 Johns Hopkins University and NBER, Baltimore, USA

2 University of Chicago and NBER, Chicago, USA

2014). The economic burden of cigarette smoking is thought to be enormous, and it is a burden that motivates public health policies targeted at smoking. For example, Xu et al. (2015, 2021) estimate that 8.7% of United States health care spending in 2010 was due to cigarette smoking and that 60% of this cost was paid by public programs such as Medicare and Medicaid. These estimates are similar to a large body of work on the social costs of smoking, which are often cited in benefit-cost analyses of tobacco policies and in civil litigation such as the 1998 Master Settlement Agreement (Manning et al., 1991; Hodgson, 1992; Cutler et al., 2000). Yet results in these highly influential studies were estimated using cross sectional data—comparing health care utilization and expenditures of smokers to never smokers at a point in time. The cross-sectional approach does not address the non-random nature of smoking status and, specifically, how health and smoking are dynamically related over the life course. Thus, current estimates of the economic costs of smoking are likely to be significantly biased, perhaps by an order of magnitude (Darden et al., 2018).

In this article, we develop a novel theoretical model of smoking and health care expenditures that is rooted in the Grossman (1972) tradition, and we use the model to guide an empirical analysis of the effect of smoking on healthcare expenditures. The model highlights the non-random nature of smoking; how health, health care expenditures, and smoking evolve with age; and the value of using longitudinal information to identify the causal effect of smoking on medical care expenditures. According to the model, smoking generates two fundamental dynamic selection problems. First, smoking causes poor health and increases the probability of death; second, poor health causes smoking cessation.[1] Both of these model implications imply that cross-sectional comparisons at a given age are comparisons of never smokers and *relatively healthy* smokers, which suggests that such comparisons may underestimate the economic costs of smoking. Yet, because smoking causes excess mortality, the expected value of the discounted sum of total medical expenditures over many years may be more or less for smokers.

Empirically, we test the predictions of our model for a sample of persons age 65 and older using a unique data merge. We link information about participants of the National Health Interview Survey (NHIS) to longitudinal data from Medicare claims.[2] These data provide information about current and retrospective smoking status, socioeconomic characteristics, and extensive, high-quality longitudinal information on a person's expenditure on medical care. With these data, we can describe, in detail, the age pattern of the use of medical care by smoking status (current, former, never) for up to 22 years. Importantly, the data allow us to gauge the role of selection by comparing, for example, the medical care utilization of smokers at age 66 to the medical care utilization of smokers at age 75, *when they were 66*. In this case, our model predicts the latter group to be healthier

[1] Economists have provided much of the evidence that health shocks can cause smokers to quit (Khwaja et al., 2006; Arcidiacono et al., 2007; Darden, 2017).

[2] Decker et al. (2012), who study medical care utilization among the previously uninsured, is the only other paper of which we are aware to exploit the Medicare claims/NHIS linkage.

than the former group and to have lower expenditures at age 66 because the latter group survives to age 75. Because we only observe smoking behavior once, our estimates measure the differences in the age profiles of medical expenditures by smoking status (current, former, and never smokers) at the age of NHIS response.

Results of our analysis are consistent with the theoretical model and lead to several novel findings. First, between ages 66 and 84, cross-sectional comparisons of current, former, and never smokers show that, *at a given age*, former smokers generate between $1,100 and $1,800 more in annual expenditures than current and never smokers, for whom there is little difference. In contrast to our cross-sectional results, results from our longitudinal analysis suggest that, prospectively, smokers generate significantly more expenditures conditional upon being alive, but significantly less when summed from ages 65 to 84. For example, smokers at age 65 who survive to age 75 incur $4,393.25/year in excess expenditures at age 75 relative to never smokers at age 65. However, at age 65, the expected value of the discounted sum of total of expenditures from ages 65 to 84 is $503.14 *less* ($124,216.03 vs. $124,719.20) for smokers vs. never smokers, which reflects significant excess mortality. Importantly, the expected value of the discounted sum of total of expenditures is less for smokers regardless of survey age (i.e., the age at which smoking behavior is measured). Conditional on living to age 75, the expected value of the discounted sum of total expenditures is $16,354.83 less for smokers relative to never smokers. Furthermore, we find that, prospectively, excess expenditures among smokers are driven largely by inpatient expenditures on the extensive margin of inpatient expenditures - smokers consistently experience inpatient stays at higher rates than never smokers. Interestingly, retrospective comparisons of smokers and never smokers show that smokers generate significantly *less* in expenditures largely due to lower outpatient, elective expenditures among smokers. This is not true for former smokers; for example, retrospective annual expenditures at age 69 are $1,100 lower for age 75 smokers but they are $589 *higher* for age 75 former smokers. To summarize, we find that smokers are more expensive in any given year, but because of excess, smoking-attributable mortality, we find no evidence that smoking is a financial burden on Medicare.

Our results inform a significant literature on the lifetime social costs associated with smoking, which have been heavily debated in the context of public policy and litigation. For example, Manning et al. (1989) and Manning et al. (1991) find that cigarette smoking generates excess medical expenditures of roughly $ 0.26/pack ($0.67/pack in 2022) when averaged over smokers of all ages. In spite of this finding, those authors argue that, because smoking causes reductions in expected longevity, smoking “pays for itself” over the life course through significantly lower retirement and nursing home expenditures. Viscusi (2002) revisits these estimates and, in particular, adjusts for changes in the likely health (mortality) effects of cigarettes associated with changes in the amount of tar in cigarettes. Viscusi (2002) concludes that cigarettes pay for themselves while still generating $0.58/pack ($1.08/pack in 2022) in extra medical expenditures. Our contribution to this literature is to demonstrate that smoking actually generates *savings* on medical care expenditures within the Medicare population. That is, conditional on living to age 65, smokers generate lower Medicare expenditures relative to nonsmokers over their time in Medicare.

We also contribute to the literature on the effects of smoking on Medicaid expenditures. For example, Cutler et al. (2000) describe cross-sectional methods that their team used in the 1990s to study the effect of smoking on expenditures in the Massachusetts Medicaid program, estimates of which were influential in 1998 Master Settlement Agreement in which the tobacco industry settled with 46 states that brought suit to recover excess medical expenditures incurred by public payers. They found that in both physician and inpatient services, former smokers generate the highest expenditures. Our cross-sectional results are similar; at a given age, former smokers generate between $1,100 and $1,800 more in annual expenditures than current and never smokers, for whom there is little difference. This result highlights a key prediction from our model that cessation may be preceded by poor health. Cutler et al. (2000) also reported that current smokers have modestly greater healthcare spending than never smokers mainly due to costly inpatient stays. These estimates are likely downward biased because the unmeasured health of smokers is likely better than never smokers. In fact, when we condition on the sub-sample of individuals not observed to die during the period of analysis (i.e. survivors) - as may be the case in a Medicaid population of younger smokers, we observe small to no difference in healthcare expenditures between current smokers and never smokers. These findings demonstrate the importance of unmeasured health and the likely bias of estimates in Cutler et al. (2000). However, Cutler et al. (2000) do not address mortality and they do not measure the difference in the lifetime Medicaid costs of smoking (current or former smokers). We document that cumulative Medicare healthcare expenditures, which are affected by mortality and the number of years of spending, are no different between smokers and never smokers. For any given birth cohort, our results imply that lifetime Medicaid expenditures would be similar if not lower for smokers despite being greater in any given year. While Cutler et al. (2000) study a younger sample, much of the greater costs of smoking they report is associated with long-term care and inpatient care among older persons in their sample, so our results for an older cohort of persons are likely relevant, but some caution is warranted given the difference samples, objectives and, approaches. Our work emphasizes the importance of longitudinal data in the context older smokers who experience significant excess mortality.

Our results have several health and economic implications. The Food and Drug Administration (FDA), which has regulatory authority over cigarettes, conducts benefit-cost analyses on proposed cigarette restrictions and on the value of harm reduction related to nicotine-delivery products. The magnitude of the health care cost of smoking is a critical input in FDA cost-benefit analyses. For example, consider the benefit-cost analysis of the recent FDA proposal to ban menthol-flavored cigarettes.[3] To the extent that such a ban reduces the prevalence of cigarette smoking, the benefit-cost analysis of this regulation requires an understanding of the extent to which changes in traditional smoking due to these regulations will cause changes in future health expenditures (Meltzer, 1997). Our results suggest significantly

[3] See https://www.fda.gov/news-events/fda-voices/fda-track-take-actions-address-tobacco-related-health-disparities

higher expenditures relative to cross-sectional comparisons. As a result, using cross-sectional estimates will understate the excess healthcare costs associated with additional cigarette smoking among older Americans. However, our results also suggest that lifetime healthcare costs of smoking may be relatively small and that the healthcare benefits (savings) of reductions in smoking may be relatively small too. Results from our study also inform many other benefit-costs analyses, including raising the legal purchase age of cigarettes to 21 (Ahmad, 2005); the benefits of early childhood interventions that affect smoking (Belfield et al., 2006); and the voluminous literature on increases in state and local cigarette taxes (Ahmad & Franz, 2008; Congressional Budget Office, 2012).

Our results also inform health policy beyond Medicare. For example, the Affordable Care Act allows insurers to charge smokers up to a 50% surcharge on insurance plans and the surcharge is not eligible for federal subsidies (Kaplan et al., 2014). Because the major health implications of smoking are not realized until an individual's 50s and 60s (Doll et al., 2004; Darden et al., 2018), our results suggest there is little justification for this surcharge for younger smokers - smokers not at risk of death are, if anything, significantly cheaper because of lower outpatient utilization. Furthermore, several studies have shown that surcharges in health insurance premiums for smokers have led to fewer smokers being covered, and smokers, who are of lower socioeconomic status, are an already vulnerable group. An arguably better policy would be to encourage smokers to obtain insurance and make use of subsidized medical services (e.g., physician counseling) that encourage cessation prior to the major health implications of smoking. Even amongst the Medicare population, our results suggest that the *overall* health care expenditure implications of smoking are economically insignificant.

2 Background

Adult smoking prevalence in the United States has declined from 42.4% in 1965 to 16.8% in 2014.[4] The reduction in cigarette smoking has been attributed to a combination of higher taxes (Chaloupka & Warner, 2000; Chaloupka et al., 2012), public health information campaigns (Smith et al., 2001; Sloan, 2003), and indoor smoking bans (Carton et al., 2016). Despite the marked decline, in 2014, roughly one out of six Americans (16.8%) smoked cigarettes, and the rate of decline in smoking has dropped. Furthermore, data from the CDC (2015) show that there are great disparities in smoking by gender (18.8% males/ 14.8% female), ethnicity (e.g., 29.2% of Native Americans), education (e.g., 43% of GED graduates), poverty status (e.g., 26.3% in poverty), and region (e.g., 20.7% in the Midwest). This socioeconomic and demographic heterogeneity in rates of smoking is important when considering the health care cost implications of smoking because it demonstrates that smokers and nonsmokers are not necessarily alike. Smokers are poorer and less educated than never smokers, and poverty and education are also highly correlated with health care

[4] https://www.cdc.gov/tobacco/data_statistics/tables/trends/cig_smoking/index.htm

use. Thus, absent effects of smoking, we expect health care expenditures to differ between smokers and never smokers

Cigarette smoking is the single greatest preventable risk factor for mortality and morbidity. Tobacco smoke contains more than 7,000 chemicals which quickly travel from the lungs into the blood stream. Cigarette smoking has been biologically linked to cancers of the bladder, cervix, esophagus, kidney, larynx, lung, mouth, pancreas, and stomach. Indeed, 9 out of 10 men who die from lung cancer have a smoking history. Cigarette smoke increases both heart rate and blood pressure because the nicotine in tobacco smoke activates the sympathetic nervous system (HHS, 2010). Not surprisingly, there exists a causal relationship between smoking and cardiovascular disease (both directly and indirectly through the effect of smoking on health markers such as blood pressure (HHS, 2010)), and several reproductive maladies. Nearly one-fifth of the annual deaths (440,000) are attributed to smoking in the United States. The 2004 United States Surgeon General report (HHS, 2004) on smoking concludes: "Smoking harms nearly every organ of the body, causing many diseases and reducing the health of smokers in general."

The substantial and severe adverse effects of smoking on health suggest that smokers will, while alive, use greater amounts of health care and incur greater health care expenditures than never smokers. However, as documented, smokers and never smokers differ by socioeconomic status and other attributes that cause smokers and never smokers to may have different health seeing behavior that would affect medical expenditures, for example, decreasing expenditures if smokers are less likely to seek care. Smokers also may die earlier potentially reducing lifetime medical expenditures and offsetting the effect of worse health. Thus, lifetime costs attributable to smoking are uncertain, as is the age profile of the health care costs of smoking, and estimating the excess medical costs of smoking may vary greatly depending on how the issues just described are addressed.

Hodgson (1992) is the seminal source for most health care cost projections with respect to smoking. His study used cross-sections of the population in seven age groups from NHANES, NHIS, National Nursing Home Survey (NNHS), and American Cancer Prevention Study II and Medicare files to simulate "longitudinal profiles" of smokers and never smokers. He reported that, among males, lifetime medical costs of heavy smokers are 47 percent higher than for never smokers. Hodgson (1992) findings concur with those of Manning et al. (1989) that cigarette smoking causes a lifetime increase in medical costs. The Hodgson (1992) study and studies similar to it are problematic for a number of reasons. First, it assumed that the cross-sectional outcomes of older individuals, for example, a male age 60, reflect the expected future outcomes of younger individuals, for example, a male age 40. This assumption ignores differences in expenditures due to birth-cohort effects, survival probabilities, and smoking prevalence that are likely to be significant. For example, life expectancy has been increasing over time. Second, survival probabilities were adjusted by only age, sex, and smoking status and not for documented differences in socioeconomic status and other factors that are correlated with smoking and survival. Third, the study used the average charge per hospital day to measure hospital costs, but this overstates costs significantly, because hospital charges have been shown to be well above hospital costs. Finally, data in Hodgson (1992) are from the

1970s and 1980s, when smoking prevalence was significantly higher with less socioeconomic differences between smokers and never smokers.[5]

In a more recent study, the Congressional Budget Office (2012) examined the effect of raising the excise tax on cigarettes on health, health care expenditures, and the federal budget. They used data from the 2000-2008 Medical Expenditure Panel Survey linked to data from the 1998-2007 National Health Interview Survey, which has smoking duration information. Their cross-sectional analysis controls for sex, race/ethnicity, education, marital status, income, geographic location, alcohol consumption, categories of body mass index health insurance coverage, and attitudes toward risk taking (as measured by the receipt of flu shots, seat belt use, likelihood of taking more risks than the average person, and belief in one's ability to overcome illness without medical help). Estimates indicated that smoking raises annual health care expenditures in the US by between 6 and 14 percent.

To summarize, it is clear that smoking causes poor health. Our point is that estimates produced by past research are likely to be biased and potentially off by an order of magnitude. We argue that two important sources of bias have been ignored:

1. Selective Mortality: Smoking causes excess mortality. Relative to never smokers, more smokers die at a given age. The implication of this fact is that surviving smokers are healthier than all smokers in a given birth cohort and by *more so* than are surviving never smokers relative to all never smokers.
2. Nonrandom Smoking Cessation: There is strong evidence that smoking cessation is not random. For example, Khwaja et al. (2006), Arcidiacono et al. (2007), and Darden (2017) show that older smokers quit smoking following a cardiovascular event or cancer diagnosis. Furthermore, Fishman et al. (2003) report that health care costs for former smokers are higher in the year of smoking cessation.

Darden (2017) and Darden et al. (2018) demonstrate the potential magnitude of the bias due to ignoring dynamics. Using data from the decades-long Framingham Heart Study and a dynamic model that allows smoking behavior (quitting) to respond to health shocks, Darden and colleagues show that the effect of lifelong smoking on expected longevity is likely to be overstated by as much as 50% (relative to lifelong nonsmokers, 4.3 years versus 9.3 years).

As our brief review highlights, there remains much uncertainty about the costs of smoking on health care expenditures. Previous study has lacked a theoretical framework and relied on cross-sectional methods that are likely biased. We address these issues in this article and provide important evidence to inform both theory and policy about the health care costs of smoking. The importance of obtaining better

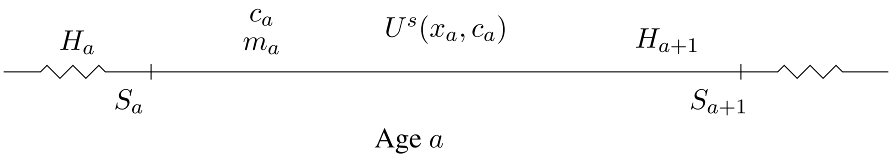
[5] Another study that simulated "longitudinal profiles" is Ahmad and Franz (2008), who used data from the Behavioral Risk Factor Surveillance Survey (BRFSS) to estimate the effect of an increase in cigarette taxes on health outcomes, tax revenue, and medical care expenditures. Importantly, they calibrated their model with health projections from Hodgson (1992), and, thus, this study has the same limitations just described. They found that raising cigarette taxes by 40% in 2004 would reduce smoking prevalence to 15.2% in 2025, increase tax revenue by $365 billion over that span and reduce total smoking-related medical costs by $317 billion.

estimates of the effect of smoking on health care costs is underscored by the wide variety of public and private decisions that depend on such estimates. We emphasize that our results are relevant to any cost-benefit analysis of tobacco control policy.

3 Theory

To understand the relationship between smoking and healthcare expenditures, and how this relationship evolves with age, we develop a dynamic model of medical care utilization and cigarette smoking. The model is of a single cohort as they age. Individuals choose smoking behavior and outpatient medical care to maximize present discounted utility, which depends on consumption (x), cigarette smoking (c), and health. The relevant state variable is an individual's stock of health, which is influenced by smoking, medical care, and age. The stock of health determines the need for future inpatient care and whether an individual dies. Whereas smoking may generate contemporaneous utility in a consumption sense, investment in costly outpatient medical care is valuable because it slows the depreciation of health. The dynamic nature of the model is helpful to show how the underlying health compositions of smokers and nonsmokers evolve, and it highlights the empirical problems with cross-sectional comparisons of outcomes by smoking status at a given age. The model generates testable hypotheses that we take to data in subsequent sections.

Consider a cohort of individuals at age a. The timing of the model proceeds according to the following timeline:



Just prior to the beginning of age a, an individual's health, H_a, evolves according to the following law of motion:

$$H_a = \Big(1 - d(c_{a-1}, m_{a-1}, a-1)\Big) H_{a-1} + \gamma(H_{a-1}) I(H_{a-1}) 1[H_{a-1} < \overline{H}] + \epsilon_a. \quad (1)$$

Health at age a is function of lagged health, which depreciates at a rate that varies by smoking behavior c_{a-1}, outpatient medical care m_{a-1}, and age (Grossman, 1972). We assume that depreciation is increasing in smoking and age (i.e., $d_c, d_{a-1} > 0$) and decreasing in medical care (i.e., $d_m < 0$). If health at age $a-1$ is below threshold $\overline{H}$, then inpatient medical care is required at age $a-1$. We assume that the amount of inpatient medical care is decreasing in the health stock (i.e, $I'(H_{a-1}) < 0$), and we assume that the marginal product of inpatient care $\gamma(H_{a-1})$ varies by the stock of health. The evolution of underlying health includes a stochastic component ϵ, which we assume follows a normal distribution: $\epsilon_a \sim N(0, \sigma^2)$. Once ϵ_a, and thus H_a, are realized, the level of H_a determines whether and the extent to which an individual requires inpatient medical care $I(H_a)$ and whether an individual dies. We categorize these health states with the variable S_a as follows:

$$S_a = \begin{cases} 2 \text{ if } \overline{H} < H_a \\ 1 \text{ if } 0 < H_a < \overline{H} \\ 0 \text{ if } H_a < 0 \end{cases} \tag{2}$$

Health category two represents good health, where no inpatient care is required (i.e., $I(H_a) = 0$). Threshold $\overline{H}$ represents the minimum level of health for which no inpatient medical care is required, and for $\overline{H} > H_a$, an increasing amount of inpatient care is used as H_a declines. Finally, health category zero represents death at age a. We separate the notions of H_a and S_a to allow for both endogenously chosen outpatient medical care, which improves health, and inpatient medical care that is not directly chosen but is influenced by smoking and other factors through effects of these factors on health. Furthermore, factors that make health uncertain also generate uncertainty in health category S_a.

Conditional on survival to age a, an individual selects the level of cigarette consumption c_a and the level of outpatient medical care m_a to maximize contemporaneous utility plus the expected present discounted value of future utility. Contemporaneous utility, $U^s(x_a, c_a)$, is a function of composite outside consumption x_a and cigarette consumption c_a, where we assume that utility is increasing at a decreasing rate in each argument. The overall health category s indexes utility to allow for variation in the marginal utility of consumption by health status and because we normalize the utility of death to be zero (Darden, 2017; Viscusi & Evans, 1990). Indexing utility by health state also allows for health to be valued for consumption, where psychological costs associated with inpatient medical care may generate incentives for health investment. As we study a relatively elderly population, we abstract from any labor market or savings/borrowing decisions and assume a simple budget constraint in which general consumption is determined by fixed, annual income W_a, such that $x_a = W_a - p_c c_a - p_m m_a - p_I I(H_a)$, where p_c, p_m, and p_I represent the exogenous and time-invariant prices of cigarettes, outpatient medical care, and inpatient medical care, respectively. At the end of age a, an individual's health category S is determined by the realized value of H_{a+1}.

3.1 Optimal behavior

Entering age a, the value of health level H_a and health category s is represented by the familiar Bellman Equation:

$$V^s(H_a, a) = \max_{c_a, m_a} \left[U^s(W_a - p_c c_a - p_m m_a - p_I I(H_a), c_a) + \beta \left(\sum_{j=1}^{2} P(S_{a+1} = j) EV^j(H_{a+1}, a+1) \right) \right], \tag{3}$$

where we substitute the budget constraint for composite consumption X_a. The value of being in health state H_a and health category s at age a is given as the level of cigarette consumption c_a and medical care m_a that maximizes current utility plus the expected present discounted value of future utility. When selecting c_a and m_a, individuals must forecast the extent to which these choices will impact H_{a+1}, which influences the probability of each health category S. Furthermore, the expectation operator is taken over the distribution of ϵ because the level of future health determines the level of future inpatient medical care, which draws from future consumption. That is, when an individual selects c_a and m_a, the level of inpatient care at age a is pre-determined because underlying health H_a is pre-determined; however, c_a and m_a will affect the probability of being in the bad health state.

Differentiating Eq. 3 with respect to cigarette consumption and rearranging yields:

$$p_c + \beta \frac{d_c}{U_x^s}\Big[\sum_{j=1}^{2}\Big(\frac{\partial P(S_{a+1}=j)}{\partial H_{a+1}} EV^j(H_{a+1}, a+1) + P(S_{a+1}=j)\frac{\partial EV^j(H_{a+1}, a+1)}{\partial H_{a+1}}\Big)\Big] = \frac{U_c^s}{U_x^s}, \tag{4}$$

which states that the optimal level of smoking is where the marginal cost of smoking (left-hand side) is equal to the marginal benefit of smoking (right-hand side). The marginal benefit is just the marginal utility of contemporaneous consumption of cigarettes. The marginal cost is the loss of general consumption due to expenditure on cigarettes plus the fact that smoking causes future health (H_{a+1}) to depreciate by d_c. The reduction in underlying health due to smoking has two effects. First, for a given expected value of a health category, it changes the corresponding probability of realizing that value; second, for a given probability of each health category, it changes the corresponding expected value. For the good health state ($S_{a+1} = 2$), the change in the expected value is just through a reduction in H_{a+1}; for the bad health state ($S_{a+1} = 1$), the change in the expected value is also through the loss of consumption because inpatient care increases as health decreases.

Similarly, differentiating Eq. 3 with respect to medical care and rearranging, we have:

$$p_m = \beta \frac{d_m}{U_x^s}\Big[\sum_{j=1}^{2}\Big(\frac{\partial P(S_{a+1}=j)}{\partial H_{a+1}} EV^j(H_{a+1}, a+1) + P(S_{a+1}=j)\frac{\partial EV^j(H_{a+1}, a+1)}{\partial H_{a+1}}\Big)\Big], \tag{5}$$

which implies that the optimal level of outpatient medical care is where the marginal benefit equals the marginal cost. The marginal cost of care is just the foregone utility from consumption. The marginal benefit depends on the value of the change in health associated with medical care (d_m). Medical care m_a is costly in that it draws from current consumption, but it generates value because underlying health H_{a+1} increases the probability of avoiding both the bad health category (and inpatient care) and death.

The model is helpful in clarifying channels through which smoking and medical care, both inpatient and outpatient, will move together or apart. The effect of smoking on inpatient care is positive unambiguously. Smoking lowers health and increases the probability of requiring inpatient care. Before discussing the direct causal pathways linking smoking to outpatient care, we highlight two exogenous factors that will generate a negative correlation between smoking and outpatient medical care. First, those who heavily discount the future will be more likely to smoke, because a small value of β lowers the marginal cost of smoking, and less likely to use outpatient medical care, because a small value of β lowers the marginal benefit of care.[6] Second, lower income individuals will be more likely to smoke and less likely to use medical care. The expected value of life, $EV^j(H_{a+1}, a+1)$, is lower for those with lower incomes because of lower consumption, X_{a+1}, which decreases the incentive to seek medical care and increases the incentive to smoke (by reducing the marginal cost of smoking), all else equal. Low-income individuals will use less care because the marginal cost of care is higher (i.e., larger U_x).[7] As income and time preference are fixed in our model, differences in smoking and outpatient medical care that result from these factors appear at any age, which highlights important time invariant determinants of smoking and medical care. More directly, smoking adversely affects health, which, as shown in Eq. 5, lowers the value of medical care because it reduces the expected value of future utility. However, a lower level of health and probability of being in the good health state is likely to raise the productivity of medical care with respect to these components of future utility. Therefore, the relationship between smoking and outpatient care is theoretically ambiguous.

3.2 Dynamic selection

At age a, total medical care expenditure is the sum of pre-determined inpatient care $I(H_a)$ and chosen outpatient care m_a, and the relevant difference is between this sum for smokers and never smokers at a. Our theory demonstrates how cross-sectional comparisons of total medical care expenditures at age a suffers from two related dynamic selection problems. The implication of both problems is that comparisons of medical expenditures at age a are likely to be between never smokers and *relatively healthy* smokers.

The first dynamic selection problem stems from non-random smoking cessation. A long literature has demonstrated that a key reason smokers quit is because of poor health (Darden, 2017; Arcidiacono et al., 2007; Khwaja et al., 2006). In our model, smoking causes health and health affects the marginal costs and benefits of smoking. For example, if the probability of good health $S_{a+1} = 2$ is decreasing in health, then a given reduction in health due to smoking may increase the marginal cost of smoking more so for individuals with low levels of health H_a, causing them to quit. The implication of non-random cessation is that, in the cross-section at age a, *former* smokers may have larger medical care expenditures at age a relative to smokers because former smokers have realized a smoking-related health shock whereas smokers at age a have not. To the extent that this shock is persistent, current smokers may be healthier than former smokers.

The second dynamic selection problems relates to selective mortality. Equation 4 indicates that smokers at age a must have a large enough marginal utility from smoking to offset the loss of consumption and the health consequences of smoking.

[6] See Scharff and Viscusi (2011) for evidence that smokers have higher implied rates of time preference.

[7] The marginal cost of smoking may be larger for those with lower income because the marginal utility of consumption is higher.

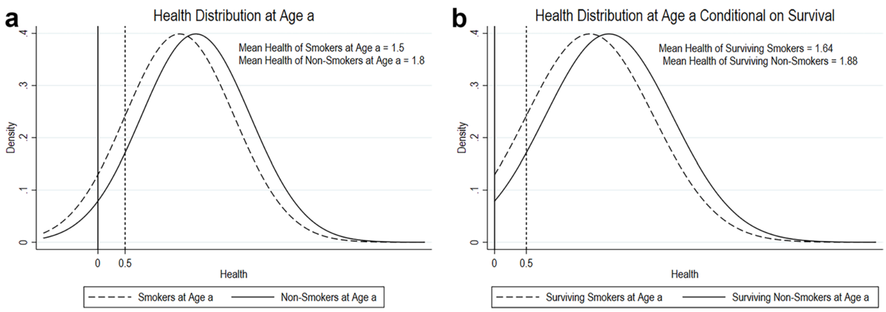


Fig. 1 Simulated Selective Mortality

A large marginal utility of smoking likely stems from an addictive reinforcement mechanism (Becker & Murphy, 1988). Because nearly all smokers start smoking in their teens, relatively older smokers at age a were almost certainly smokers (or of some significant smoking history) at age $a-1$. Equation 1 shows that smoking at $a-1$ causes a reduction in H_a through the depreciation of health, which implies, all else equal, that the distribution of H_a for smokers at $a-1$ is to the left of that of never smokers at $a-1$. As a result, more smokers die at the beginning of age a. For both smokers and never smokers at age $a-1$, mortality at the beginning of age a causes the observed distribution of health H_a to be truncated at zero, the implication of which is that the survivors to age a where healthier, on average, than the entire population at age a. Our main point is that this difference is larger for smokers at age a because smoking at age $a-1$ causes excess mortality. To see this, the truncated mean of health is:

$$E\left(H_a | H_a > 0\right) = \mu_a + \frac{\sigma \phi(\frac{-\mu_a}{\sigma})}{1 - \Phi(\frac{-\mu_a}{\sigma})}, \tag{6}$$

where $\mu_a = \left(1 - d(c_{a-1}, m_{a-1}, a-1)\right)H_{a-1} + \gamma(H_{a-1})I(H_{a-1})1[H_{a-1} < \overline{H}]$. For age $a-1$ smokers, μ_a is smaller because $d_c > 0$. However, this implies that the update factor (i.e., $\frac{\phi(.)}{1-\Phi(.)}$) is larger for age $a-1$ smokers - smoking may cause the mean of health of smokers to decline, but relative to never smokers, mortality of smokers has a larger selection effect. The magnitude of this selection effect depends on how smoking affects the hazard of death. Using very different estimators, both Doll et al. (2004) and Darden et al. (2018) show almost no difference in the longevity curves of lifelong smokers and never smokers up to age 50, at which point the curves deviate significantly. Furthermore, notice that the update factors both converge to zero in age, which suggests that the implications of selective mortality decline at older ages, when a significant fraction of the cohort has died.

To demonstrate the importance of selective mortality, Fig. 1 presents the distribution of health at different ages by smoking status while assuming that the threshold for inpatient care is 0.5 (i.e., $\overline{H} = 0.5$) and the threshold for death is $H = 0$. Figure 1a shows the distribution of underlying health, H_a, for smokers and never smokers just before the health category evolves from S_{a-1} to S_a, that is, prior to the realization of ϵ_a. According to the health evolution equation in Eq. 1, health is distributed normally, and because smoking causes health to depreciate faster, the distribution for smokers (dotted line) is to the left of that for never smokers (solid line). The mean of underlying health is 1.5 for smokers and 1.8 for never smokers. Figure 1b shows the distribution of health after the realization of ϵ_a, and thus, S_a. Smokers are more likely to die, and while the mean of health of surviving smokers increases by 0.14 relative to at a, the similar increase for non-smokers is only 0.08.

The example in Fig. 1 demonstrates that selective mortality will bias estimates of the effect of smoking on medical care expenditure to the extent that smoking increases the hazard of mortality.

3.3 Summary and testable hypotheses

Our theory implies that dynamic selection, both due to non-random cessation and selective mortality, causes cross-sectional comparisons of health and medical care expenditures at a given age to be biased downwards. At young ages, dynamic selection will be less important as differential mortality is negligible (Doll et al., 2004; Darden et al., 2018) and because non-chronic health depreciation (e.g., worsening cardiovascular biomarkers) are not typically associated with cessation (Darden, 2017). However, as a cohort ages into their fifties and sixties, we expect dynamic selection to grow in importance to some age, which implies that our cross-sectional results with respect to Medicare beneficiaries should be smaller than cross-sectional results from younger ages. Aside from dynamic selection, our theory highlights the endogeneity of smoking with respect to expenditures. Smoking behavior is chosen jointly with outpatient medical care, and failing to control for heterogeneity with respect to time preferences and socioeconomic characteristics will bias comparisons. We take the following hypotheses to data:

1. Conditioning on survival at a given age, the difference in medical expenditures between smokers and never smokers from that age should be smaller retrospectively than prospectively.
2. Conditioning on survival at a given age, the prospective difference in medical expenditures between smokers and never smokers should be *larger* in longitudinal data than in cross-sectional data.
3. Conditioning on survival at a given age, prospective inpatient medical expenditures should be larger for smokers.
4. Conditioning on survival at a given age, prospective outpatient medical expenditures may be larger or smaller for smokers.
5. Conditioning on survival at a given age, prospective and retrospective medical expenditures may be larger for former smokers than current smokers if poor health causes former smokers to quit.

Both with respect to dynamic selection and endogeneity, our theory emphasizes the importance of longitudinal data. None of the above hypotheses are testable without longitudinal information on expenditures. As we discuss below, our data allow us to construct longitudinal profiles of both inpatient and outpatient medical expenditures merged to a snapshot of smoking behavior. These data allow us to test the implications of our theory.

4 Data and descriptive analysis

Our data represent a unique linkage of survey information and medical claims made available by the National Center for Health Statistics (NCHS). As the foundation of our data, we start with repeated cross-sections of the National Health Interview Survey (NHIS) from 1997 through 2013. NHIS data provide information about a respondent's tobacco use, socioeconomic characteristics (e.g., education) and other information at a specific point in time (i.e., year). For example, the 1998 NHIS collects information from respondents about whether they ever smoked (100 cigarettes in lifetime); age first smoked regularly; number of days smoked in past 30 days; number of cigarettes smoked per day; and, for those who have stopped smoking, the time since quitting. To these data, we merge Medicare Part A, B, and D claims data from two files: the Summary Medicare Enrollment and Claims (SMEC) file from 1991 through 1998 and the Master Beneficiary Summary File Cost and Use Segment (MBSF) from 1999 through 2013. We also merge year and age of death information from the National Death Index Linked Mortality File. The resulting data allow us to follow NHIS respondents over time, both retrospectively (backward in time) and prospectively (forward in time) using the Medicare claims data to measure their use of and expenditures on medical care.

Because we study smoking in the Medicare population, we select all persons in NHIS between the ages 65 and 86 at the time of the survey with non-missing responses to the questions regarding past and current smoking behavior.[8] Table 1 provides examples of the variation generated by our data linkage for NHIS survey years 1999 and 2007 and for a variety of survey ages. For example, the 2007 NHIS survey contains information on individuals between the ages of 65 and 86. For a 70-year old in the 2007 NHIS sample, we observe their Medicare claims back to 2002 when they were 65 years of age. Prospectively, we observe Medicare claims on this person through 2013, when they are 76 years of age. If the person dies between age 70 and 76, we observe this through the mortality file. For this person, we observe 11 years of longitudinal claims information, but in some cases, we observe individuals for up to 21 years. Our final sample, comprised of respondents from each NHIS survey from 1997-2013, consists of 54,600 unique individuals and 651,733 person/year observations.

[8] Most individuals start Medicare at age 65. We drop individuals for whom Medicare claims appear prior to age 65. Ages in NHIS are topcoded at 86.

Table 1 Demonstration of Longitudinal Data

Survey/Year	Age at Time of Survey	Year First Observed in Claims	Age First Observed in Claims	Year Last Observed or Year Age 86	Age Last Observed in Claims	Number of Years Observed
NHIS 1997	65	1997	65	2013	81	16
	70	1994	65	2013	86	21
	75	1991	69	2008	86	17
	80	1991	74	2003	86	12
NHIS 2007	65	2007	65	2013	71	6
	70	2002	65	2013	76	11
	75	1999	67	2013	81	14
	80	1999	72	2013	86	12

The Table demonstrates the variation we use to identify our model. Age varies within NHIS surveys, which vary by year. The Table shows example contributions to our sample of individuals at difference ages who took their NHIS survey in 1999 and 2007. NHIS surveys from 1999-2013 were linked to claims information from 1999-2013. NHIS surveys in 1997 and 1998 were linked to Medicare claims from 1991-2013

Table 2 provides cross-sectional summary statistics of our sample by NHIS-reported smoking behavior. At the NHIS survey, 48.3% of respondents reported having not smoked more than 100 cigarettes in their lifetimes, 40.6% reported past smoking, and 11.1% reported being current smokers, which we define as daily smoking. Those reporting smoking at the NHIS survey may be different in than never smokers in other ways that drive costs. For example, NHIS summary statistics in Table 2 suggest that, relative to never smokers, current smokers are more likely to be, younger, male, African-Americans with lower education, income, and health. Our theory demonstrates that elective medical care and cigarette smoking are, in part, determined by similar time-invariant heterogeneity, which highlights the importance of longitudinal data. Table 2 also includes the fraction in each smoking category that we observe to die in the National Death Index. Current smokers and former smokers are 12.1 and 3.6 percentage points, respectively, more likely to die during our sample period than never smokers.

We define variables for total expenditures, total inpatient expenditures, total outpatient expenditures, physician expenditures, and skilled nursing facility expenditures. Table 3 lists the source file and input variables used to create each expenditure dependent variable. From 1999-2013, we use the MBSF file to construct total expenditures as the sum of all inpatient expenditures (acute and other) and all outpatient, prescription drug, skilled nursing facility, and hospice expenditures. From the MBSF file, we define these variables as the sum of MBSF variables that capture a.) Medicare expenditures b.) other, 3rd party expenditures and c.) out-of-pocket cost-sharing expenditures. Outpatient expenditure variables include claims related to several cost/use categories, including anesthesia, ambulatory surgical center, physician, evaluation and management, hospital, dialysis, imagining, testing, durable medical equipment, and other expenditures. Outpatient expenditures also include Part B and

Table 2 Summary Statistics of NHIS Variables

	Never Smokers	Former Smokers	Current Smokers
Proportion	0.483	0.406	0.111
Age at Survey	75.314	74.367	71.707
	(0.040)	(0.042)	(0.070)
Female = 1	0.725	0.435	0.535
	(0.003)	(0.003)	(0.006)
Black = 1	0.117	0.101	0.157
	(0.002)	(0.002)	(0.005)
Asian = 1	0.032	0.015	0.012
	(0.001)	(0.001)	(0.001)
Hispanic = 1	0.100	0.065	0.073
	(0.002)	(0.002)	(0.003)
High School = 1	0.311	0.295	0.317
	(0.003)	(0.003)	(0.006)
Some College = 1	0.202	0.231	0.222
	(0.002)	(0.003)	(0.005)
College Degree or More = 1	0.184	0.191	0.101
	(0.002)	(0.003)	(0.004)
Education Variable Missing = 1	0.005	0.004	0.005
	(0.000)	(0.000)	(0.001)
Currently Working	0.115	0.122	0.128
	(0.002)	(0.002)	(0.004)
Low Income = 1	0.612	0.560	0.703
	(0.003)	(0.003)	(0.006)
Owns a Home = 1	0.761	0.785	0.673
	(0.003)	(0.003)	(0.006)
Medicaid = 1	0.117	0.086	0.156
	(0.002)	(0.002)	(0.005)
Sample Weight	5778.725	6275.551	5487.207
	(23.777)	(26.744)	(47.794)
Poor Health = 1	0.241	0.260	0.309
	(0.003)	(0.003)	(0.006)
Observed to Die	0.357	0.393	0.478
	(0.003)	(0.003)	(0.006)
N	26,374	22,185	6,041

The table presents summary statistics, means and proportions, of NHIS variables by NHIS smoking behavior. n=54,600

Part D expenditures. We construct separate dependent variables for physician expenditures and for skilled nursing facility expenditures.

As shown in Table 1, we observe expenditures for up to 21 years on individuals between the ages of 65 and 86, and our data contain significant variation in the age of interview, the age of claims, and the age of death. We structure our data such

Table 3 Medical Expenditure Variables

Variable	Source File	Variable Label
Total expenditures	SMEC 1991-1998	Total Medicare Expenditures
	MBSF 1999-2013	Acute Inpatient expenditures + Other Inpatient expenditures +
		Anesthesia + Ambulatory Surgery Center +
		Dialysis + Durable Medical Equipment +
		Eval. and Management + Home Health +
		Hospice + Hospital Outpatient +
		Imaging + Other Procedure +
		Physician + Part B Drug +
		Part D Drug + Skilled Nursing Facility +
		Test expenditures
Total Inpatient expenditures	SMEC 1991-1998	Inpatient Total Expenditures
	MBSF 1999-2013	Acute Inpatient expenditures +
		Other Inpatient expenditures
Total Outpatient expenditures	SMEC 1991-1998	Outpatient Total Expenditures
	MBSF1999-2013	Anesthesia + Ambulatory Surgery Center +
		Dialysis + Durable Medical Equipment +
		Eval. and Management + Home Health +
		Hospital Outpatient + Imaging +
		Other Procedure + Physician +
		Part B Drug + Part D Drug +
		Test expenditures

Data were made available through a secure linkage program through the National Center for Health Statistics. Summary Medicare Enrollment and Claims (SMEC) data include aggregated categories above. Medicare Beneficiary Summary File Cost and Use Segment contains the disaggregated claims types, which we use to construct aggregated measures

that claims age represents time, and our goal is to compare expenditure outcomes across different NHIS survey ages and Medicare claims ages. To maximize statistical power in our empirical work below, we define three-year age bins as follows:

$$A^r_{ia} = \begin{cases} 1 \text{ if } a \in [65, 67]; \text{ denoted age } 66 \\ 2 \text{ if } a \in [68, 70]; \text{ denoted age } 69 \\ 3 \text{ if } a \in [71, 73]; \text{ denoted age } 72 \\ 4 \text{ if } a \in [74, 76]; \text{ denoted age } 75 \\ 5 \text{ if } a \in [77, 79]; \text{ denoted age } 78 \\ 6 \text{ if } a \in [80, 82]; \text{ denoted age } 81 \\ 7 \text{ if } a > 82; \text{ denoted age } 84, \end{cases} \tag{7}$$

where $r \in \{C, S\}$ for claims and survey ages. Note that while A^c_{ia} varies within individual, A^s_i is claims age invariant - an individual participates in NHIS only once.

Using this notation, Fig. 2 demonstrates the claims age profile of total expenditures for three different NHIS survey ages (66, 72, and 78) for those who had never

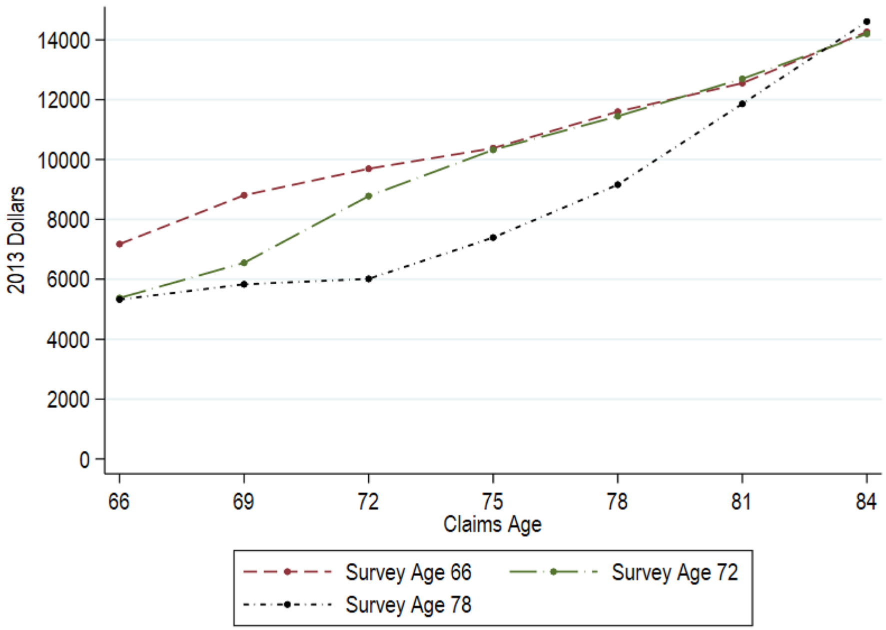


Fig. 2 Mean Annual Expenditures for never smokers at Three Survey Ages

smoked at their NHIS interview. The x-axis of Fig. 2 represents claims age, where we denote each claims age by the mid-point of three-year age bins (e.g., age 66 for the age bin [65, 67]). Thus, mean expenditures by claims age represent the average annual expenditure within a given age range. Notice that at claims age 66, there exists significant variation in total Medicare expenditures by NHIS survey age. For example, for those never smokers for whom we observe their NHIS survey at age 66, real (2013 $) total expenditures at age 66 are roughly $2,000 higher than those never smokers whom we observe in NHIS at ages 72 and 78. This reflects selective mortality amongst *never smokers*. By conditioning on living to age 78, for example, evidence in Fig. 2 demonstrates that the survey age 78 group are healthier on average than the survey age 66 group *at age 66*. Furthermore, this gap persists through claims age 81. In the next Section, our goal is to understand how this selection varies by smoking behavior.

5 Econometric specification

Our empirical strategy is motivated by the theoretical model above. In that model, we have two broad types of health care expenditures: outpatient and inpatient care. The model predicts that the amount of each type of care will evolve with age because of differences in the evolution of health with age. Smoking affects the evolution of health, and thus the amount of inpatient and outpatient care. The different evolution of health between smokers and never smokers will cause a divergence in the use of various types of healthcare with age between smokers and never smokers. Thus, our empirical strategy leverages differences in expenditures within survey age categories across smoking status. Define m_{iat} as a measure of medical expenditures for individual i at age a in year t. The main regression model is:

$$\begin{aligned} m_{iat} = & \sum_{j=1}^{7}\sum_{k=1}^{7} \beta_{jk} 1[A_i^s = j]1[A_{ia}^c = k] + \sum_{j=1}^{7}\sum_{k=1}^{7} \gamma_{jk} 1[A_i^s = j]1[A_{ia}^c = k]1[\text{ Former Smoker}_i = 1] + \\ & + \sum_{j=1}^{7}\sum_{k=1}^{7} \delta_{jk} 1[A_i^s = j]1[A_{ia}^c = k]1[\text{ Current Smoker}_i = 1] + \sum_{l=1991}^{2013} \lambda_t 1[t = l] + x_i\psi + \epsilon_{iat}. \end{aligned} \tag{8}$$

Equation 8 allows the claims age profiles of expenditures, indexed here by k, to vary by NHIS survey age category, indexed here by j, and by NHIS-reported smoking status. The intercept of the model, β_{11} represents the mean, annual medical expenditure for individuals between ages 65 and 67 who responded to the NHIS survey between ages 65 and 67 and who report never smoking. The β parameters capture the mean, annual expenditure age profile for never smokers by NHIS survey age category, and the γ and δ parameters capture how the claims age profiles differ for former and current smokers, respectively. The vector x_i includes individual demographic and socioeconomic characteristics listed in Table 2, and the λ parameters capture claims year-specific effects.

Equation 8 demonstrates the importance of longitudinal data. In the absence of longitudinal data, Equ. 8 would be restricted such that $\beta_{jk} = \gamma_{jk} = \delta_{jk} = 0\ \forall j \neq k$. In other words, a cross-sectional analysis assumes that medical expenditures of smokers at any claims age are the same regardless of their survey age. So, a smoker observed at age 66 would have the same medical expenditures at age 66 as a smoker observed at age 76 (if the expenditure of the 76 year old could be measured retrospectively). This is unlikely to be the case because the pool of smokers who are alive at age 76 are likely healthier than the pool of smokers alive at age 66. Accordingly, they are likely to have lower expenditures at age 66 because of their better health. The same age-related bias applies to never smokers and former smokers. In contrast, the longitudinal data that we use allows us to hold constant the survival age and compare people with the same survival age (health). Moreover, we can measure medical expenditures prospectively and retrospectively. The prospective differences in medical expenditure between smokers and never smokers (former smokers) are more accurate measures of the differences in medical expenditure between smokers and never smokers (former smokers).

To demonstrate the role of selective mortality, and thus the changing composition of smokers relative to never smokers, we take two approaches. First, we compare our estimates of parameters in Eq. 8 to those from a sub-sample for whom we do not observe death. From a given survey age, the difference in δ parameter estimates between these two samples sheds light on the role of smoking on expenditures by, in the case of survivors, holding the sample composition fixed. Second, we estimate a linear probability model for mortality that follows from Eq. 8. Estimates of β and δ from this model allow us to construct the hazard of death for never smokers and current smokers, which we use to construct the expectation of excess expenditures by survey age.

Our data allow us to reasonably test the implications of selective mortality with respect to expenditures, but prospective expenditure (mortality) comparisons are still affected by non-random smoking status and likely differences in unmeasured health at a given age. To address this problem, we re-estimate the model (i.e., Eq. 8) with interactions between baseline (at time of NHIS survey) controls, such as race/ethnicity, gender, self-reported health and education, and cohort age dummies. These baseline factors are all well-known correlates of health. The motivation for this specification is that the relatively large set of interactions plausibly control for time(age)-varying differences in unmeasured health that may confound the longitudinal comparisons between smokers and never smokers. We present these results below, but note here, that estimates are virtually unchanged when these interactions are included in the model. While not definitive evidence that there is no remaining confounding, we view this finding as substantial support that our estimates are reasonable estimates of the causal effect of smoking.

5.1 Inpatient vs. Outpatient expenditures

The wealth of claims information allows us to estimate versions of Eq. 8 for a wide variety expenditure measures. Differences in the effects of smoking on, for example, outpatient compared to inpatient expenditures shed light on health-seeking behavior of smokers motivated in Sect. 3. Furthermore, because the distributions of many of these expenditure variables are right-skewed, and because, even in the Medicare population, many exhibit significant point mass at zero, we also estimate two-part models for each outcome, following from the long literature on modeling medical expenditures (Deb et al., 2017). In particular, we model the extensive and intensive margins of inpatient expenditures separately because the intensive margin asks whether, conditional on positive inpatient expenditures, those expenditures differ for smokers vs. never smokers while the extensive margin asks whether smokers and never smokers have different rates of inpatient utilization.

6 Results

Figure 3 presents regression adjusted total expenditures from Eq. 8 in which we restrict $\beta_{jk} = \gamma_{jk} = \delta_{jk} = 0\ \forall j \neq k$. The resulting cross-sectional differences are as if there was not a longitudinal component to our data. At survey age 66, current smokers (black line) have the highest expenditures ($8,455.76 per enrollee/year), followed by former smokers ($8,305.54 per enrollee/year) and never smokers ($7,172.68). However, at all ages through 84, *former* smokers have the highest total annual expenditures whereas the never and current smoking trends are indistinguishable.

Projecting medical care expenditures from, say, age 66 using results in Fig. 3 yields the result that current smokers are expected to have the same expenditures as never smokers. We contrast these results with results from estimating Eq. 8 without the coefficient restrictions. Table 4 presents the estimated γ and δ parameters for each survey age (column). Two important themes emerge from the results in Table 4.

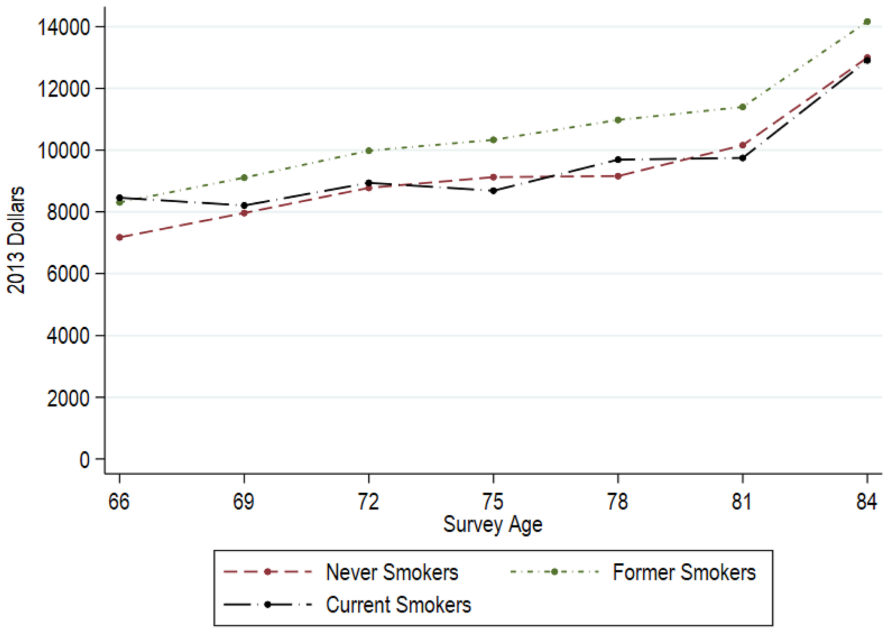


Fig. 3 Cross-Sectional Age Profile by Smoking Status: $\beta_{jk} = \gamma_{jk} = \delta_{jk} = 0\ \forall j \neq k$

First, moving down each column, the longitudinal results show that total annual expenditures are positive and significant *after* the claims age passes the survey age. That is, looking prospectively from the survey age, when smoking is measured, the δ parameter estimates reveal that smokers incur significantly higher expenditures. For example, current smokers between the ages of 71 and 73 incur higher total annual expenditures at ages 75 ($3,826.32), 78 ($4,289.51), 81 ($4,564.38) and 84 ($ 5,308.82). The second theme that emerges is that, prior to survey age, current smokers incur expenditures that are significantly less than never smokers. At these claims ages, we know that individuals are not at risk of death because they appear later to take the NHIS survey, and uniformly the estimates are either not statistically significant or significantly negative. For example, the current smoker at age 72 incurred $ 616.87 and $515.41 less than never smokers at ages 69 and 66, respectively.

The adverse effects of smoking (measured prospectively) are particularly large three to five years after the person is observed in the survey, but this is only the case for those observed prior to age 75 or so. At age 75 and beyond, smoking does not have much effect on expenditures despite the fact that smoking still seems to worsen health, as evidenced by higher rate of dying among smokers from age 75 onward. It is also the case that once that jump in healthcare expenditures occurs, expenditures tend to remain high, but not continue to diverge. These results suggest that smoking causes a transition to the bad health state—in terms of our theory—and the person stays in that state until death.

The bottom panel of Table 4 presents results for former smokers relative to never smokers. Prospectively, the pattern in results is similar to current smokers: former

Table 4 Main Regression Estimates for Total Medicare expenditures

	Panel A: Current Smokers vs. Never Smokers: δ_{jk}						
	NHIS Survey Age Bin: A_i^s						
Claims Age Bin: A_{ia}^c	[65, 67]	[68, 70]	[71, 73]	[74, 76]	[77, 79]	[80, 82]	≥ 83
[65, 67]	1283.08***	-938.85**	-554.41	-605.60	-474.55	-167.06	
	(495.17)	(354.45)	(387.92)	(509.25)	(775.56)	(2449.14)	
[68, 70]	1220.67**	242.20	-616.87*	-1100.88***	-599.56	85.36	-2522.02**
	(514.65)	(412.52)	(370.27)	(429.43)	(736.53)	(1448.72)	(1067.26)
[71, 73]	3187.64***	1284.51**	162.61	-960.80*	-1034.27**	284.17	-1388.84**
	(789.38)	(627.61)	(517.27)	(542.66)	(500.40)	(996.40)	(649.65)
[74, 76]	4393.25***	2266.47***	3826.32***	-440.56	-849.30*	-1602.49***	-1520.03**
	(899.55)	(765.10)	(873.63)	(562.21)	(515.35)	(598.66)	(669.23)
[77, 79]	5315.75***	2530.95***	4289.51***	847.81	532.41	-1082.29	-863.96
	(1469.88)	(924.83)	(1070.83)	(716.81)	(623.59)	(669.37)	(696.75)
[80, 82]	3217.98**	3525.44**	4564.38***	232.91	981.51	-412.75	-458.77
	(1593.17)	(1434.52)	(1344.01)	(991.07)	(934.36)	(841.80)	(662.83)
≥ 83	2751.46	4519.72*	5308.82**	172.19	329.65	1174.86	-91.28
	(6197.24)	(2394.79)	(2145.78)	(1230.64)	(1219.21)	(1175.23)	(650.56)
	Panel B: Former Smokers vs. Never Smokers: γ_{jk}						
	NHIS Survey Age Bin: A_i^s						
Claims Age Bin: A_{ia}^c	[65, 67]	[68, 70]	[71, 73]	[74, 76]	[77, 79]	[80, 82]	≥ 83
[65, 67]	1132.86***	110.07	774.67***	74.68	-136.56	-2025.15***	.
	(309.50)	(287.78)	(267.54)	(297.40)	(385.52)	(800.91)	.
[68, 70]	1560.86***	1144.96***	1233.14***	589.28**	216.69	-32.47	399.53
	(392.91)	(345.32)	(285.93)	(284.35)	(354.94)	(426.20)	(1108.83)
[71, 73]	1877.95***	1116.12**	1204.46***	690.13**	704.54***	530.14	-62.27

Table 4 (continued)

	Panel A: Current Smokers vs. Never Smokers: δ_{jk}						
	NHIS Survey Age Bin: A_i^s						
	(514.80)	(437.85)	(356.24)	(303.13)	(273.18)	(337.10)	(335.58
[74, 76]	2125.21***	1969.91***	1859.08***	1208.51***	1242.96***	647.46	604.20**
	(625.13)	(530.42)	(443.09)	(368.27)	(308.88)	(333.64)	(285.27)
[77, 79]	1329.98*	2078.91***	1788.99***	1381.83***	1819.12***	732.05	1038.23***
	(723.60)	(656.59)	(545.71)	(481.55)	(373.12)	(359.45)	(271.15)
[80, 82]	1561.34	1556.63*	1338.27**	2223.30***	1868.61***	1239.72	853.96***
	(1254.05)	(890.43)	(672.05)	(613.07)	(533.36)	(423.21)	(268.02)
≥ 83	5995.22	-96.66	1685.15*	794.42	955.11*	1560.36	1169.58***
	(4593.13)	(1125.07)	(880.14)	(650.88)	(556.20)	(504.48)	(296.60)
Mean Annual expenditures at Survey Age: Never Smokers	7172.68	7962.99	8776.76	9124.76	9158.20	10159.12	12999.19

Table presents estimates of γ and δ from Eq. 8. The dependent variable is total Medicare expenditures. Standard errors are clustered at the individual level. Estimates are conditional on NHIS variables included in Table 2

* p-value < 0.1. n=651,733; ** p-value < 0.05; *** p-value < 0.01

smokers generate higher Medicare expenditures prospectively following the NHIS survey. However, especially at younger survey ages, expenditures are larger for former smokers relative to never smokers prior to the relevant survey age. Former smokers may have already experienced a smoking related health shock which both caused them to quit and increased expenditures. We view these results as evidence of non-random cessation. Consistent with this hypothesis – recent health issue causes cessation – at older survey ages, retrospective expenditure differences between former smokers and never smokers are smaller further back in time. Yet to appear in our data, former smokers must have survived any health shocks, which suggest that these shocks were comparatively mild. To investigate, Table 5 presents estimates of γ and δ parameters from linear probability models for mortality. For each survey age, we estimate the how the probability of death varies by smoker status from that survey age until age 84. For all survey ages, both current and former smokers face a higher probability of death between the survey age and age 84. Former smokers exhibit smaller excess mortality relative to current smokers, consistent with results in Table 4 that prospective excess expenditures are smaller for former smokers, and consistent with the hypothesis that mortality attributable to smoking is what drives excess expenditures.

To make the comparison with the cross-sectional estimates, Fig. 4 presents the estimated δ parameters for survey ages 66 (panel a.), 72 (panel b.), and 78 (panel c.). In each panel, the left figure presents the longitudinal parameter (δ) estimates (red line) along side the cross-sectional (black line) parameters (γ) that are constant across survey ages. The right figure of each panel presents estimates from longitudinal and cross-sectional linear probability models with death as the dependent variable. By definition, the red-line in the right figures starts at the relevant survey age because the probability of death prior to the survey age is zero. At each survey age, Fig. 4 demonstrates significant differences between the longitudinal and cross-sectional estimates. For example, at survey age 72, excess expenditures were negative for smokers prior to age 72 but significantly positive after age 72 (in contrast with cross-sectional results). The right figure for survey age 72 reveals significant excess mortality following age 72 (again in contrast with the cross-sectional results).

Results from Table 4 and Fig. 4 strongly suggest that what makes smoking expensive with respect to medical expenditure is that smokers die prematurely and death associated with substantial expenditures. Figure 5 presents estimates of δ for three survey ages (66, 72, and 78) for both our full sample and a sub-sample for whom we do not observe death. In this sub-sample. prospective comparisons from say, age 72, are free of the selective mortality effect. At survey age 66, prospective comparisons from age 66 are significantly attenuated in the sub-sample of survivors and not significantly different than zero until age 75. At survey ages 72 and 78, δ estimates of the sub-sample of survivors are statistical zeros when looking prospectively.

In sum, when projecting excess expenditures from a given age, our results show that cross-sectional comparisons will badly misrepresent the age profile of expenditures by smoking behavior, and we provide strong evidence that selective mortality is to blame.

A concern with estimates from Eq. 8 is that individual-level unobserved heterogeneity, such as rate of time preference, income and underlying health, will

Table 5 Excess Mortality Estimates for Current Smokers

	Panel A: Current Smokers vs. Never Smokers: δ_{jk}						
	NHIS Survey Age Bin: A_i^s						
Claims Age Bin: A_{ia}^c	[65, 67]	[68, 70]	[71, 73]	[74, 76]	[77, 79]	[80, 82]	≥ 83
[65, 67]	0.002						
	(0.003)						
[68, 70]	0.022***	0.006*					
	(0.003)	(0.003)					
[71, 73]	0.031***	0.028***	0.008**				
	(0.004)	(0.004)	(0.004)				
[74, 76]	0.048***	0.036***	0.039***	0.018***			
	(0.004)	(0.004)	(0.039)	(0.005)			
[77, 79]	0.043***	0.043***	0.054***	0.050***	0.022***		
	(0.006)	(0.005)	(0.054)	(0.006)	(0.006)		
[80, 82]	0.095***	0.055***	0.071***	0.051***	0.057***	0.022**	
	(0.009)	(0.007)	(0.071	(0.007)	(0.007)	(0.009)	
≥ 83	0.061	0.044***	0.072***	0.048***	0.043***	0.054***	0.029**
	(0.043)	(0.011)	(0.072)	(0.007)	(0.007)	(0.008)	(0.007)
	Panel B: Former Smokers vs. Never Smokers: γ_{jk}						
	NHIS Survey Age Bin: A_i^s						
Claims Age Bin: A_{ia}^c	[65, 67]	[68, 70]	[71, 73]	[74, 76]	[77, 79]	[80, 82]	≥ 83
[65, 67]	-0.001						
	(0.002)						
[68, 70]	0.006***	0.001					
	(0.002)	(0.002)					
[71, 73]	0.007**	0.012***	0.000				
	(0.003)	(0.003)	(0.003)				
[74, 76]	0.008***	0.016***	0.011***	0.003			
	(0.003)	(0.003)	(0.003)	(0.003)			
[77, 79]	0.010**	0.009***	0.016***	0.018***	0.003		
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)		
[80, 82]	0.009	0.014***	0.012***	0.030***	0.021***	0.002	
	(0.007)	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)	
≥ 83	-0.003	0.015**	0.016***	0.014***	0.016***	0.019***	0.007***
	(0.027)	(0.007)	(0.005)	(0.004)	(0.004)	(0.004)	(0.003)

Table presents estimates of γ and δ from Eq. 8. The dependent variable is total Medicare expenditures. Standard errors are clustered at the individual level. Estimates are conditional on NHIS variables included in Table 2

* p-value < 0.1.; ** p-value < 0.05; *** p-value < 0.01

jointly influence smoking behavior and health expenditures. To assess the likely importance of this potential confounding, we re-estimated Eq. 8, but included a large set of interactions between observed baseline characteristics and cohort age dummy variables. Figure 6 presents estimates from this model and, for

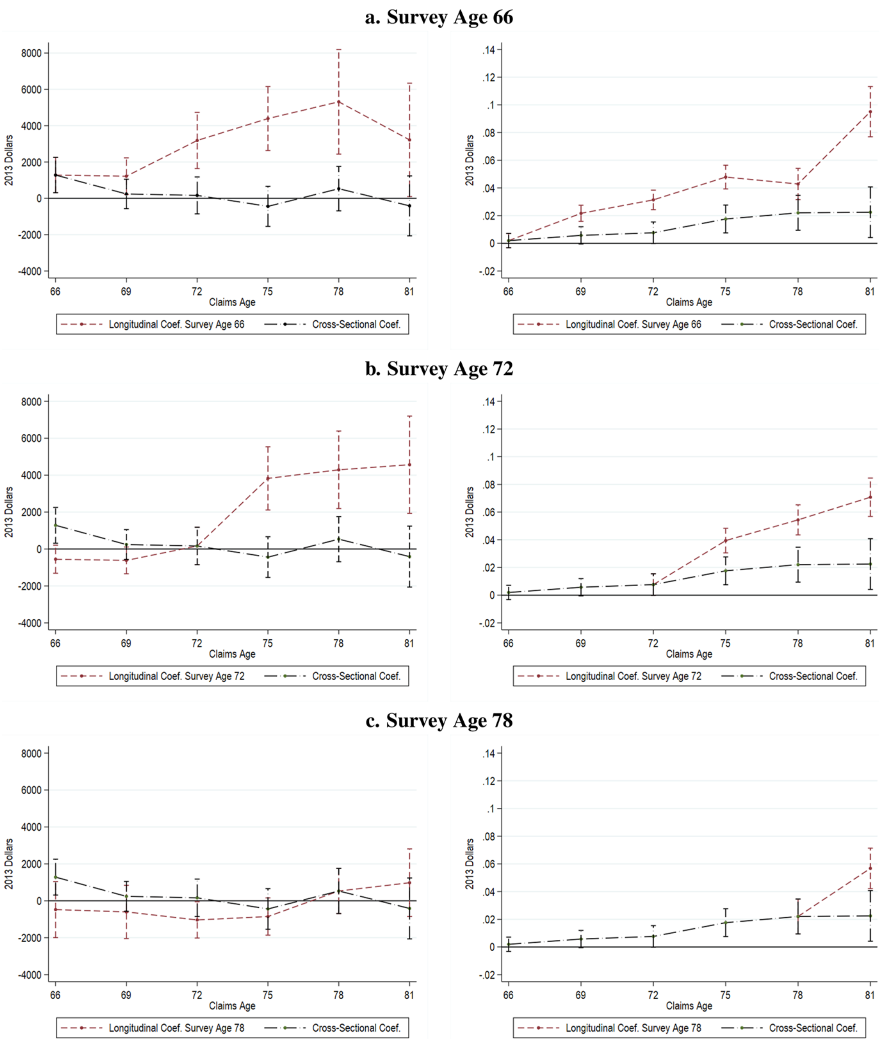
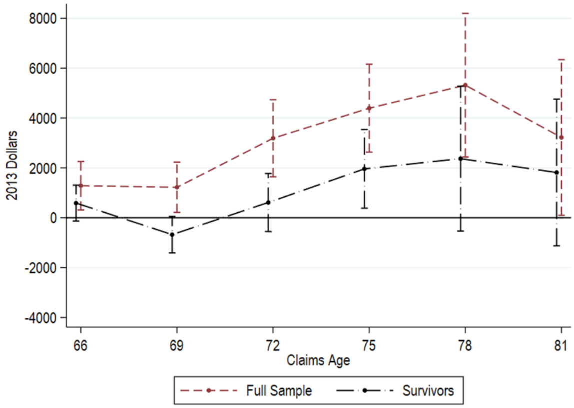
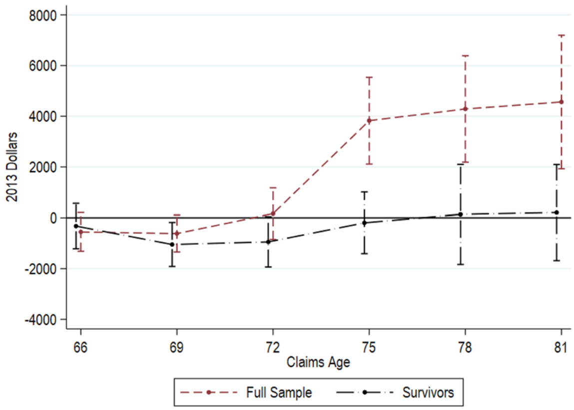


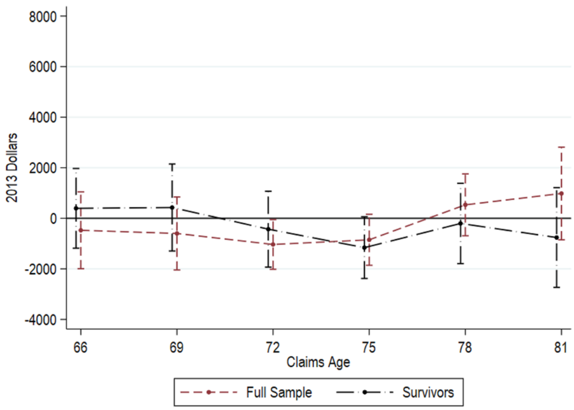
Fig. 4 Excess Expenditures (Left) and Mortality (Right) for Current Smokers relative to Never Smokers

comparison, estimates from Table 4. The main point to note about this figure is that at each survey age, estimates from both models are nearly identical. There is little evidence of (time-varying) selection on observable variables that are known correlates of health, which makes it plausible that there is relatively little selection on unmeasured factors. Holding constant the age of survival, as we did in the longitudinal analysis, is a powerful way to control for selection. Finally, the *differences* in estimates in Table 4 at adjacent claims age years within a survey age provide another, although imperfect, control for time-invariant factors that may cause differences in health care expenditures between smokers and never smokers. Although there is some compositional change and selective mortality

Fig. 5 Estimates of δ for Full Sample vs. Survivor Sub-Sample

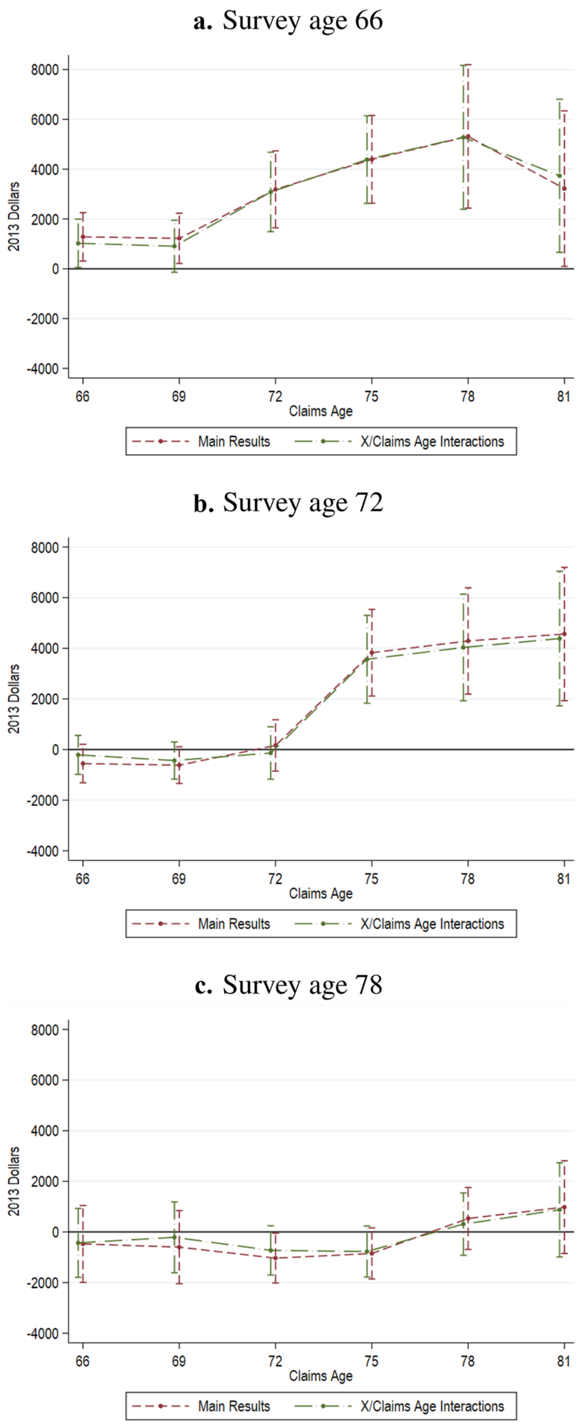






between adjacent claims ages, the samples are roughly equivalent and, therefore, differences in healthcare expenditures between the two ages hold constant the influence of time-invariant factors. Looking at the difference (in difference) in expenditures of smokers versus never-smokers suggests a similar conclusion

Fig. 6 Estimates of δ Allowing Claims Age Profiles to Vary by X_i



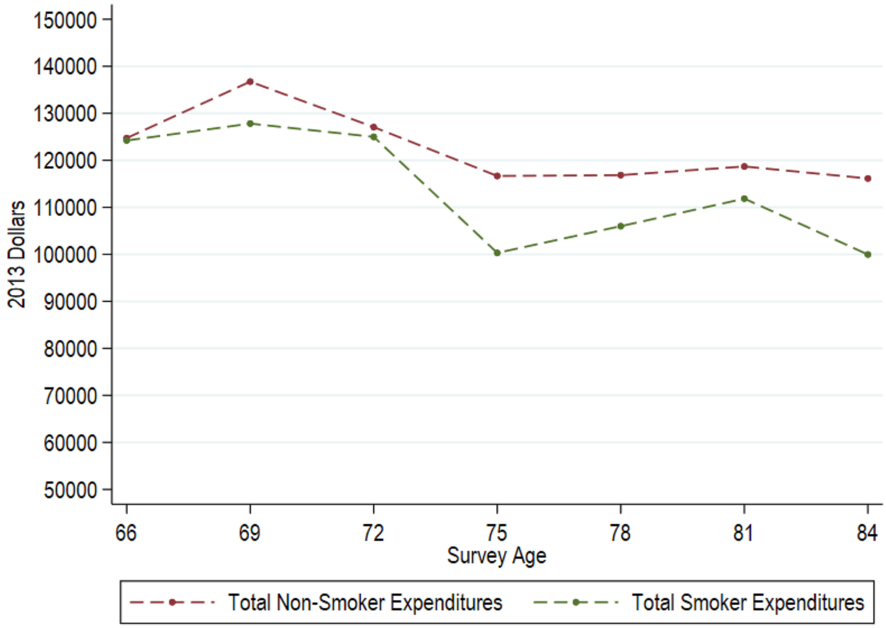


Fig. 7 Expected Excess Expenditures Summed from Age 66 to 84 and discounted to Age 65 by Survey Age

— smoking worsens health and causes an increase in expenditures while alive and higher rate of mortality.

Next, we aggregate our results from Eq. 8 by survey year. To do so, we use estimates from the linear probability model of mortality to construct the survival rates at each age. That is, prior to a survey age, the survival rate is equal to one (hazard of death = 0), but following the survey age, the hazard of death rises and the survival rate declines. We use the survival rate to construct the expected value of the sum of medical expenditures from age 66 to age 84 for each survey age. We discount this sum back to age 65 for all survey ages, which is important because by not smoking, never smokers are delaying expensive end-of-life care. We assume a discount rate of 3% per year. As an example of these calculations, consider a person who we observe at age 75 in the NHIS survey. For this person, the survival rate is equal to one between ages 66 and 75 and declines after age 75. So, the sum of expenditures form age 66 to 84 will include expenditures measured retrospectively and prospectively. We then discount this expected value back to age 65.

The results are presented in Fig. 7. For those observed in the NHIS survey at age 66, the expected total of expenditures from ages 66 to 84 is \$503.14 lower for smokers than never smokers; this difference is \$16,354.83 when comparing smokers and never smokers at age 75. For all survey ages, smokers are more expensive while alive, as shown in Fig. 4, but the significant excess mortality relative never smokers implies that the expected value of the discounted sum of expenditures is lower. Figure 7 shows that, at all survey ages, the expected value of the discounted sum of total of expenditures is lower for smokers. Consistent with our theory of expenditures generally declining with age at survey, those who survive to older ages are healthier and their expenditures are lower retrospectively. The decline of the expected total sum of expenditures is on average larger for smokers than never smokers, which is also consistent with the relatively stronger selection on health among smokers.

To investigate the sources of the expenditure differences, we examined inpatient and outpatient expenditure models separately. Table 6 presents estimates total inpatient expenditures (top panel) and outpatient expenditures (bottom panel). For inpatient expenditures, we find very similar trends in excess expenditures as we saw in Table 4. Prospectively from a given survey age, inpatient expenditures are significantly positive for smokers relative to never smokers; retrospectively, however, estimates of the differences in expenditures between smokers and never smokers are not statistically different than zero (and small in magnitude). On the other hand, outpatient expenditures are more consistently negative prior to survey age, which is consistent with the theoretical prediction that, for smokers in relatively good health, outpatient expenditures may be lower because of unobservable factors that drive the decision to continue smoking and explain lower levels of health-seeking behavior. Outpatient expenditures are largely significantly positive following the survey age because of the complementarity between inpatient and outpatient medical care for relatively sick individuals.

Similarly, we estimate two-part models based on Eq. 8 with respect to inpatient expenditures. The top panel of Table 7 presents estimates for the intensive margin (log of inpatient expenditures), and the bottom panel presents estimates from a linear probability model of any inpatient expenditures. Looking prospectively, inpatient expenditures are significantly more expensive for current smokers relative to never smokers at younger survey ages (i.e., survey ages 66, 69, and 72), but not significantly different at older survey ages. In the bottom panel of Table 7, extensive margin estimates of δ show that smokers experience inpatient care significantly more than never smokers across all survey ages. Importantly, we find no significant differences in inpatient expenditures on the extensive margin when looking retrospectively – those claiming to smoke at the NHIS survey have yet to experience the major health implications of smoking. This result also explains why we see lower total expenditures for smokers retrospectively (i.e., similar rates of inpatient stays and lower outpatient expenditures).

7 Discussion and conclusion

Results of our analysis are consistent with the theoretical predictions of our model. Smoking is a cause of poor health and greater medical expenditures. However, because smokers are selected on health, this relationship is only apparent prospectively after conditioning on age of survival, which is a powerful control for underlying and unmeasured health. Prospectively, smokers incur more healthcare expenditures because they are sicker and this is compellingly evidenced by their higher use of inpatient care and higher mortality rates. Retrospectively, healthcare expenditures of smokers are lower than never smokers and this is likely due to two reasons:

Table 6 Effects of Current Smoking on Inpatient and Outpatient Expenditures

	Inpatient Expenditures: δ_{jk}						
	NHIS Survey Age Bin: A_i^s						
Claims Age Bin: A_{ia}^c	[65, 67]	[68, 70]	[71, 73]	[74, 76]	[77, 79]	[80, 82]	≥ 83
[65, 67]	917.54***	-228.19	1.42	19.80	-129.07	241.13	.
	(267.20)	(199.48)	(218.35)	(288.05)	(377.91)	(1845.64)	.
[68, 70]	840.09***	610.00***	-12.16	-233.90	-234.32	866.62	-1462.60***
	(280.46)	(208.87)	(205.15)	(255.24)	(307.07)	(1075.50)	(546.64)
[71, 73]	2139.25***	1303.29***	501.18*	-225.55	-232.83	376.16	-521.64
	(451.15)	(367.26)	(289.25)	(295.19)	(264.94)	(503.13)	(387.97)
[74, 76]	2155.93***	1751.17***	2802.91***	357.27	-144.96	-559.49	-436.29
	(462.38)	(436.08)	(575.13)	(316.60)	(297.71)	(346.49)	(383.45)
[77, 79]	3103.80***	1393.35***	2092.12***	543.85	579.29	293.94	-293.20
	(806.99)	(484.09)	(552.49)	(376.13)	(350.43)	(427.77)	(402.43
[80, 82]	1407.13*	1581.10**	2116.20***	501.39	1009.88*	143.40	315.55
	(823.95)	(704.01)	(665.30)	(519.03)	(559.28)	(418.60)	(407.04
≥ 83	1065.55	1541.08*	2776.97**	1106.26*	563.72	911.51	-54.56
	(3643.55)	(916.30)	(1096.21)	(634.05)	(607.51)	(591.53)	(290.19)
	Outpatient Expenditures: δ_{jk}						
	NHIS Survey Age Bin: A_i^s						
Claims Age Bin: A_{ia}^c	[65, 67]	[68, 70]	[71, 73]	[74, 76]	[77, 79]	[80, 82]	≥ 83
[65, 67]	287.65	-647.65***	-286.44	-406.32	-269.85	-159.07	
	(231.21)	(188.11)	(189.98)	(277.83)	(417.37)	(804.46)	
[68, 70]	159.22	-624.07***	-428.54**	-612.95***	-125.74	-656.94	-751.62
	(247.93)	(223.60)	(179.00)	(221.07)	(481.82)	(426.44)	(701.16)
[71, 73]	488.89	-181.77	-160.46	-441.46	-562.70**	-75.96	-721.50**
	(330.72)	(285.22)	(232.05)	(273.64)	(261.58)	(484.67)	(306.30)
[74, 76]	1141.54***	-56.99	541.59*	-565.82**	-647.69**	-595.98**	-897.61***
	(369.65)	(331.08)	(321.93)	(237.13)	(220.39)	(278.98)	(291.01)
[77, 79]	841.64*	460.49	1103.01**	-330.47	-293.31	-921.29***	-593.53**
	(499.93)	(388.72)	(431.95)	(289.10)	(264.45)	(272.99)	(272.10)
[80, 82]	552.62	574.46	1165.96**	-407.36	-440.74	-603.32**	-561.11**
	(621.67)	(529.10	(542.34)	(400.17)	(317.85)	(296.34)	(235.50)
≥ 83	1637.37	686.37	573.42	-753.33	-838.19*	-460.34	-70.25
	(2048.63)	(841.44)	(704.78)	(473.84)	(436.89)	(375.88)	(262.91)

Table presents estimates of γ and δ from Eq. 8. The dependent variables are total inpatient expenditures (top panel) and total outpatient expenditures (bottom panel) as defined in Table 3. Standard errors are clustered at the individual level. Estimates are conditional on NHIS variables included in Table 2

* p-value<0.1. n=651,733; ** p-value < 0.05; *** p-value < 0.01

Table 7 Effects of Current Smoking Inpatient Expenditures: Intensive vs. Extensive Margin

	Log(Inpatient Expenditures): δ_{jk}						
	NHIS Survey Age Bin: A_i^s						
Claims Age Bin: A_{ia}^c	[65, 67]	[68, 70]	[71, 73]	[74, 76]	[77, 79]	[80, 82]	≥ 83
[65, 67]	0.144***	-0.060	0.085	0.004	-0.030	-0.346	
	(0.053)	(0.065)	(0.085)	(0.135)	(0.187)	(0.454)	
[68, 70]	0.068	0.075	0.002	-0.030	-0.035	0.154	-0.881***
	(0.055)	(0.046)	(0.060)	(0.084)	(0.120)	(0.279)	(0.154)
[71, 73]	0.261***	0.127**	0.078	0.042	0.041	-0.029	-0.272
	(0.062)	(0.053)	(0.051)	(0.079)	(0.096)	(0.138)	(0.197)
[74, 76]	0.134**	0.050	0.203***	0.050	-0.118	0.043	0.050
	(0.060)	(0.054)	(0.051)	(0.055)	(0.074)	(0.095)	(0.167)
[77, 79]	0.175**	0.027	0.068	-0.006	0.083	0.235***	-0.013
	(0.085)	(0.060)	(0.063)	(0.052)	(0.063)	(0.078)	(0.100)
[80, 82]	0.083	0.070	0.161	-0.010	0.041	0.007	0.060
	(0.107)	(0.078)	(0.068)	(0.065)	(0.061)	(0.066)	(0.078)
≥ 83	-0.174	-0.025	0.102	0.093	0.076	-0.014	0.022
	(0.544)	(0.123)	(0.098)	(0.058)	(0.060)	(0.046)	(0.037)
	Any Inpatient Expenditures: δ_{jk}						
	NHIS Survey Age Bin: A_i^s						
Claims Age Bin: A_{ia}^c	[65, 67]	[68, 70]	[71, 73]	[74, 76]	[77, 79]	[80, 82]	≥ 83
[65, 67]	0.023***	0.000	-0.004	0.006	0.007	0.030	.
	(0.007)	(0.007)	(0.009)	(0.013)	(0.021)	(0.067)	.
[68, 70]	0.030***	0.024***	0.008	-0.006	0.009	-0.006	-0.051
	(0.008)	(0.007)	(0.008)	(0.011)	(0.016)	(0.024)	(0.051)
[71, 73]	0.044***	0.030***	0.016*	-0.017*	-0.014	0.026	0.002
	(0.011)	(0.009)	(0.009)	(0.009)	(0.011)	(0.024)	(0.023)
[74,76]	0.078***	0.063***	0.068***	0.013	0.001	-0.014	-0.016
	(0.014)	(0.012)	(0.011)	(0.010)	(0.011)	(0.016)	(0.019)
[77, 79]	0.075***	0.067***	0.066***	0.043***	0.017	-0.006	-0.014
	(0.018)	(0.016)	(0.014)	(0.013)	(0.011)	(0.015)	(0.016)
[80, 82]	0.092***	0.069**	0.070***	0.038**	0.034**	0.021	0.005
	(0.031)	(0.021)	(0.020)	(0.016)	(0.016)	(0.017)	(0.015)
≥ 83	0.058	0.079**	0.088***	0.032	-0.004	0.056***	0.005
	(0.130)	(0.033)	(0.031)	(0.021)	(0.017)	(0.021)	(0.013)

Table presents estimates of γ and δ from Eq. 8. The dependent variables are the log of inpatient expenditures (top panel, n=102,923) and a binary indicator for any inpatient expenditures in a given year (bottom panel, n=651,733). Standard errors are clustered at the individual level. Estimates are conditional on NHIS variables included in Table 2

* p-value < 0.1; ** p-value < 0.05; *** p-value < 0.01

smokers are apt to have characteristics that make them use less outpatient care than never smokers, for example, higher rates of time preference, and because smokers are likely to be healthier than never smokers (and not use inpatient care) up to that age. Combining these two distinct patterns of differential healthcare expenditures and their underlying causes yields the finding that smokers incur less healthcare expenditures than never-smokers from ages 65 to 84. This result is true at every available survey age (i.e., the points for smokers are always below those of never smokers in Fig. 7). The primary explanation of this result is that smokers are more likely to die (see Fig. 4) and have less years of expenditures, which is a possibility noted by Zhorabian and Philipson (2010), but our paper is the first to show this result empirically.

Our study has three limitations that merit comment. First, the relationship between smoking status and healthcare expenditures, even conditional on survival and rich observable heterogeneity, may be confounded by unmeasured factors. We addressed this issue by including a relatively large set of observable characteristics known to be important correlates of health and healthcare expenditure, and allowed the healthcare expenditure age profiles to differ by these characteristics. As noted, this made virtually no difference (see Fig. 6). So, while we cannot rule out significant confounding, we have some substantial evidence that it may not be that important. Second, we only had access to data that measured smoking status at a point in time when a person was observed in the NHIS survey. For the ages in our sample, there are likely no new smokers, but there may be some churn in smoking status that makes the categories of smoking status we used, never, former, and current, imperfect classifications that may introduce some measurement error. However, if some smokers quit following their NHIS interview, then our prospective results are upper bounds on the effect of smoking on expenditures, and our main conclusion that smoking does not generate excess expenditures between ages 65 and 84 will likely be unaffected. Third, examining the healthcare costs of smoking for a sample ages 65 and older misses the dynamics between smoking and health (healthcare) that occur before this age. Thus, our paper provides a contrast to Xu et al. (2015, 2021), who study a sample of (mostly) younger adults.

The results of our analysis have important policy implications. Smokers do not appear to impose an externality on tax payers because of the risk pooling of Medicare. While in any given year, smokers are likely to spend more on healthcare than never smokers, smokers live fewer years and overall cost taxpayers less than never smokers. Also, considering the evidence that smoking has few adverse effects on health at younger ages, there does not seem to be a strong argument of a similar externality in insurance plans for the under 65 age group, which for example, underlie the ACA's policy to allow insurers on the healthcare exchanges charge smokers a premium. However, the most important policy impact of our study is to highlight how existing evidence of the healthcare costs of smoking are likely seriously biased. Indeed, cross-sectional analyses, such as that we provided, suggest that smokers and never smokers have similar expenditure patterns at older ages. These estimates are biased downward and understate the health and healthcare costs of smoking. Smokers do cost more than never smokers while alive, although the difference declines with age because smokers become increasingly positively selected on health (see Fig. 4). Our findings of lower healthcare costs of smoking from ages 66 to 84 are not biased, but instead, result from the higher mortality rates of smokers.

Funding This manuscript was prepared with linked Medicare claims/National Health Interview Survey data from the National Center for Health Statistics. The findings and conclusions in this paper are those of the authors and do not necessarily represent the views of the Research Data Center, the National Center for Health Statistics, or the Centers for Disease Control and Prevention. Dr. Darden acknowledges funding from a Johns Hopkins University Catalyst Award.

Declarations

Conflicts of interests/Competing interests Neither Dr. Darden nor Dr. Kaestner have any conflicts of interest or competing interests related to this manuscript.

References

Ahmad, S. (2005). The cost-effectiveness of raising the legal smoking age in california. *Medical Decision-Making*, *May June*, 330–340.
Ahmad, S., & Franz, G. (2008). Raising taxes to reduce smoking prevalence in the us: A simulation of the anticipated health and economic impacts. *Public Health, 122,* 3–10.
Arcidiacono, P., Sieg, H., & Sloan, F. A. (2007). Living rationally under the volcano? An empirical analysis of heavy drinking and smoking. *International Economic Review, 48,* 37–65.
Becker, G., & Murphy, K. (1988). A theory of rational addiction. *Journal of Political Economy, 96,* 675–700.
Belfield, C., Nores, M., Barnett, S., & Schweinhart, L. (2006). The high/scope perry preschool program cost-benefit analysis using data from the age-40 followup. *Journal of Human Resources, 41,* 162–190.
Carton, T., Darden, M., Levendis, J., Lee, S., & Ricket, I. (2016). Comprehensive indoor smoking bans and smoking prevalence: Evidence at the state level. *American Journal of Health Economics, 2,* 535–556.
CDC. (2015). *Current Cigarette Smoking Among Adult, United States, 2005–2014*. MMWR Morb Mortal Wkly Rep 44 Centers for Disease Control and Prevention.
Chaloupka, F., & Warner, K. (2000). The economics of smoking. *Handbook of Health Economics, 1,* 1539–1627.
Chaloupka, F., Yurekli, A., & Fong, G. (2012). Tobacco taxes as a tobacco control strategy. *Tobacco Control, 21,* 172–180.
Congressional Budget Office. (2012). *Raising the Excise Tax on Cigarettes: Effects on Health and the Federal Budget*. Report Congressional Budget Office.
Creamer, M. R., Wang, T. W., Babb, S., Cullen, K.A., Day, H., Willis, G., Jamal, A., & Neff, L. (2014). Tobacco product use and cessation indicators among adults - United States. *MMWR Morb Mortal Wkly Rep*.
Cutler, D., Epstein, A., Frank, R., Hartman, R., King, C., Newhouse, J., et al. (2000). How good a deal was the tobacco settlement?: Assessing payments to massachusetts. *Journal of Risk and Uncertainty, 21,* 235–261.
Darden, M. (2017). Smoking, expectations, and health: A dynamic stochastic model of lifetime smoking behavior. *Journal of Political Economy, 125,* 1485–1522.
Darden, M., Gilleskie, D., & Strumpf, K. (2018). Smoking and mortality: New evidence from a long panel. *International Economic Review, 59,* 1571–1619.
Deb, P., Norton, E., spsampsps Manning, W. (2017). *Health Econometrics Using Stata*. Stata Press.
Decker, S., Doshi, J., & Polsky, D. (2012). Health service use among the previously uninsured: Is subsidized health insurance enough? *Health Economics, 21,* 1155–1168.
Doll, R., Peto, R., Boreham, J., Gray, R., & Sutherland, I. (2004). Mortality in relation to smoking: 50 years' observations on male British doctors. *British Medical Journal, 328,* 1519–1528.
Fishman, P., Khan, Z., Thompson, E., spsampsps Curry, S. (2003). Health care costs among smokers, former smokers, and never smokers in an hmo. *Health Services Research*, *38*.
Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy, 80,* 223–255.
HHS. (2004). *The Health Consequences of Smoking: A Report of the Surgeon General.*. Report United States Department of Health and Human Services, Centers for Disease Control and Prevention,

National Center for Chronic Disease Prevention and Promotion, Office on Smoking and Health, Atlanta, GA.
HHS. (2010). *How Tobacco Smoke Causes Disease: The Biology and Behavioral Basis for Smoking-Attributable Disease*. Department of Health and Human Services: Report U.S.
HHS. (2014). *The Health Consequences of Smoking: 50 Years of Progress*. Department of Health and Human Services: A Report of the Surgeon General. Report U.S.
Hodgson, T. (1992). Cigarette smoking and lifetime medical expenditures. *The Milbank Quarterly, 70,* 81–125.
Kaplan, C., Graetz, I., & Waters, T. (2014). Most exchange plans charge lower tobacco surcharges than allowed, but many tobacco users lack affordable coverage. *Health Affairs, 33,* 1466–1473.
Khwaja, A., Sloan, F. A., & Chung, S. (2006). Learning about individual risk and the decision to smoke. *International Journal of Industrial Organization, 24,* 683–699.
Manning, W., Keeler, E., Newhouse, J., Sloss, E., & Wasserman, J. (1991). *The Costs of Poor Health Habits*. MA: MIT Press Cambridge.
Manning, W., Keeler, E., Newhouse, J., Sloss, E., & Wasserman, J. (1989). The taxes of sin do smokers and drinkers pay their way? *JAMA, 261,* 1604–1609.
Meltzer, D. (1997). Accounting for future costs in medical cost-effectiveness analysis. *Journal of Health Economics, 16,* 33–64.
Scharff, R., & Viscusi, K. W. (2011). Heterogeneous rates of time preference and the decision to smoke. *Economic Inquiry, 49,* 959–972.
Sloan, F. A., Smith, V. K., & Taylor, D. H. (2003). *The Smoking Puzzle: Information, Risk Perception, and Choice*. Harvard University Press.
Smith, V. K., Taylor, D. H., Sloan, F. A., Johnson, F. R., & Desvousges, W. H. (2001). Do Smokers Respond to Health Shocks? *Review of Economics and Statistics, 83*(4), 675–687.
Viscusi, W. K. (2002). *Smoke-Filled Rooms A Postmortem on the Tobacco Deal*. University of Chicago.
Viscusi, W. K., spsampsps Evans, W. N. (1990). Utility functions that depend on health status: Estimates and economic implications. *American Economic Review*, *80*.
Xu, X., Bishop, E., Kennedy, S., Simpson, S., & Pechacek, T. (2015). Annual healthcare spending attributable to cigarette smoking: An update. *American Jouranal of Preventive Medicine, 48,* 326–333.
Xu, X., Shrestha, S., Trivers, K., Neff, L., Armour, B., spsampsps King, B. (2021). U.S. healthcare spending attributable to cigarette smoking in 2014. *Preventive Medicine*, *150*.
Zhorabian, A., & Philipson, T. (2010). External costs of risky health behaviors associated with leading actual causes of death in the U.S.: a review of the evidence and implications for future research. *International journal of environmental research and public health, 7,* 2460–2472.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.