MEMORANDUM

No 07/2011

Quitting Behaviour of Cigarette Smokers
Are there direct effects of a screening program?

Anne Line Bretteville-Jensen
Erik Biørn
Randi Selmer

ISSN: 0809-8786

Department of Economics
University of Oslo

This series is published by the
University of Oslo
Department of Economics

P. O.Box 1095 Blindern
N-0317 OSLO Norway
Telephone: + 47 22855127
Fax: + 47 22855035
Internet: http://www.sv.uio.no/econ
e-mail: econdep@econ.uio.no

In co-operation with
The Frisch Centre for Economic Research

Gaustadalleén 21
N-0371 OSLO Norway
Telephone: +47 22 95 88 20
Fax: +47 22 95 88 25
Internet: http://www.frisch.uio.no
e-mail: frisch@frisch.uio.no

Last 10 Memoranda

No 06/11	Kari H. Eika *Near and Generous? Gift Propensity and Chosen Emotional Distance*
No 05/11	Finn R. Førsund *Weight Restrictions in DEA: Misplaced Emphasis?*
No 04/11	Vladimir Krivonozhko, Finn R. Førsund, Andrey V. Lychev *Terminal units in DEA: Definition and Determination*
No 03/11	André K. Anundsen, Tord S. H. Krogh, Ragnar Nymoen, Jon Vislie *Overdeterminacy and endogenous cycles: Trygve Haavelmo's business cycle model and its implications for monetary policy*
No 02/11	Nils Chr. Framstad *Portfolio Separation Properties of the Skew-Elliptical Distributions*
No 01/11	Karine Nyborg , Tao Zhang *Is corporate social responsibility associated with lower wages?*
No 22/10	Rebecca Graziani, Nico Keilman *The sensitivity of the Scaled Model of Error with respect to the choice of the correlation parameters: A Simulation Study*
No 21/10	Jennifer L. Castle, Jurgen A. Doornik , David F. Hendry, Ragnar Nymoen *Testing the Invariance of Expectations Models of Inflation*
No 20/10	Erik Biørn *Identifying Trend and Age Effects in Sickness Absence from Individual Data: Some Econometric Problems*
No 19/10	Michael Hoel, Svenn Jensen *Cutting Costs of Catching Carbon Intertemporal effects under imperfect climate policy*

Previous issues of the memo-series are available in a PDF® format at:
http://www.sv.uio.no/econ/forskning/publikasjoner/memorandum

Quitting behaviour of cigarette smokers.
Are there direct effects of a screening program?

Anne Line Bretteville-Jensen

Norwegian Institute for Alcohol and Drug Research (SIRUS) and
Institute of Health Management and Health Economics, University of Oslo

Erik Biørn

Department of Economics, University of Oslo

Randi Selmer

Norwegian Institute for Public Health

Abstract: This paper aims at i) providing effect estimates of a wide range of covariates and traditional policy means to increase the smoking cessation rate, ii) offering evidence on alternative interventions for health authorities, and iii) examining and comparing three groups of smokers with varying lengths of their smoking career (including one group that has smoked ≥ 25 years). All smokers have been subject to a three-wave cardiovascular screening and followed up over a maximum of 14 years. This rich panel data set has been merged with administrative registers. A flexible discrete-time duration model is used to examine the effect of 5 categories of explanatory variables: personal characteristics; indicators of addiction status; economic factors; health and health shock variables; governmental interventions. Most covariates differ across groups, but for all groups did the screening participation years stand out as important. Possible policy implications for future cessation interventions are discussed.

Keywords: Smoking behaviour. Cigarette cessation. Duration model. Quitting hazard. Long-term smokers. Health status and shocks. Screening. Policy intervention. Panel data.

JEL codes: C23, C24, C41, I12, I18

1. Introduction

General background

Cigarette smoking is singled out by the World Health Organisation (WHO) as the most important preventable cause of premature death in developed countries (WHO 2009). Reducing the cigarette consumption has been on the health authorities' agenda for decades and much effort has been devoted towards preventing young people from taking up the habit, reducing the number of cigarettes smoked by habitual smokers, and encouraging smokers to quit. The means for combating smoking include taxes on tobacco products, regulatory interventions (sale restrictions, age limits, smoke-free environments, etc.) and anti-smoking information and campaigns. The present paper focuses on factors influencing the quitting behaviour. Potential health improvements of cessation are substantial, even for persons who have been smoking for several years (USDHHS 1990; Ostbye and Taylor 2004; IARC 2007).

Measured by the substantial drop in the prevalence of daily cigarette smokers, the western anti-tobacco policy has been relatively successful. The declining trend for Norwegian men and women for the period 1973-2009 can illustrate this general pattern (Figure 1). Typically, the trends differ by gender, and the reduction in female smoking did only start one decade ago. In 1973 about 50% of Norwegian males and 30% of females smoked. Twenty years later, the gender gap was about to disappear, and in 2009 the prevalence of daily smokers had been reduced to about 20% for both genders.

Still further reduction in the cigarette consumption is an important target for the health authorities. One significant problem facing decision makers, however, is that a further expansion of many well-known and frequently used means for combating cigarette consumption may prove inefficient or have limited political appeal. Tax policy is considered the most effective governmental means for reducing tobacco consumption (Chaloupka and Warner 2000). Increased prices may reduce both the number of cigarettes consumed by smokers and the number of smokers, as supported by a huge number of empirical cigarette studies (see e.g., Gallett and List 2003; Levy et al. 2004 for overviews). There are limits to the applicability also of the price instrument, not least because tax increases always evoke political debates and because the steep social gradient in smoking can make support for additional increases even more difficult. Also, increases in cross-border shopping and illegal imports of cigarettes may undermine tax hikes when cigarette prices become higher than commonly “accepted”.

Many countries have already implemented an array of smoking regulations. Smoking is often prohibited in workplaces and public areas (including restaurants and bars); many have strict regulations on the selling and buying of tobacco products in addition to a ban on cigarette commercials. Further restrictions might be feasible (say, on smoking in open-air and in private homes), but may prove difficult to implement and control. And although anti-smoking information will still be needed to affect consumers’ understanding of risks and harms, the effect of traditional anti-smoking campaigns on *current* smokers may be limited due to the widespread knowledge that smoking can seriously damage health. Policy makers and others working for an improved “health of the nation” therefore seek new tools as well as improved knowledge about the efficiency of traditional means to combat cigarette smoking (Warner and Mendez 2010).

Setting of the study

In this paper we aim at estimating the influence of a wide range of covariates and traditional policy means on the cessation rate of daily cigarette smoking. We further aim at analysing the effect of alternative interventions for health authorities to increase this rate. One research objective is thus to examine the impact on smoking cessation of a three-wave screening program conducted during the study period. A novelty of this study is that we distinguish between three groups of smokers, according to the length of the pre-sample smoking career, for short named ST-smokers (Short-term smokers), MT-smokers (Medium-term smokers) and LT-smokers (Long-term smokers). Using the same model setup, while a priori accounting for group differences in the coefficient values, we can contrast the effects of explanatory factors on LT-smokers (career of at least 25 years) to the effects on smokers with shorter careers.

By international standards, our panel data set is fairly rich, not only in having a rather long observation window (up to14 years), but also in covering a large number of participants and incorporating five essentially different categories of potential explanatory variables: (a) personal characteristics, (b) indicators of addiction status, (c) economic factors, (d) health status and health shock variables, and (e) governmental interventions. A substantial part of the data set stems from administrative registrations, a fact that may reduce problems related to recall bias, etc. The panel design of the data set is particular, in that records from a three-wave panel (at a distance of roughly five years between the waves) are `superimposed' on annual register information. We analyse the data by employing a discrete-time duration model which also accounts for unobserved heterogeneity.

The estimated effects of most covariates differ across the three groups of smokers. The findings suggest for instance that anti-smoking policies succeeding in delaying smoking onset could increase the quitting hazard among ST-smokers, while policies that contribute to a less intensive smoking habit (lowering the number of daily cigarettes smoked) seem to pull in the same direction for both ST- and MT-smokers. The screening effect (represented by dummies for the relevant screening years) is clearly significant for all three groups of smokers, although stronger for LT-smokers than for the other two groups. As will be elaborated in the concluding section, we suggest that the screening effect may be partly explained by the fact that the smokers, by simply being invited to, participating, and receiving test results, are reminded of and confronted with their individual health risk. Hence, extended use of targeted screening programs, tobacco counselling by health care providers, etc., may prove increasingly useful for reaching this group of very experienced smokers.

Relation to literature

Empirical studies aiming at an improved understanding of quitting behaviour have either employed a discrete choice framework or used duration models (Forster and Jones 2001). Studies examining general smoking participation rates have proved less useful, however, as changes in such rates cannot distinguish between changes in the starting and the quitting rate, and factors influencing the two rates may differ (Douglas 1998). Hyland et al. (2004), Ross et al. (2005) and DeCicca et al. (2008) are recent examples of studies that treat quitting as a binary event using logistic and probit models, respectively. Also duration models have a binary outcome variable and focus on the risk of transition from one state to another (e.g., from non-smoking to smoking or from smoking to non-smoking). In addition to analysing the effect of relevant covariates (including time-varying variables like prices and income), this econometric approach particularly takes into account the duration of the smoking behaviour. Duration analyses of cigarette quitting include e.g., Douglas (1998); Tauras (1999); Tauras and Chaloupka (2001); Forster and Jones (2001); DeCicca et al. (2002); Nicolás (2002); Tauras (2004); Kidd and Hopkins (2004); Madden (2007); DeCicca and McLeod (2008). Some relevant findings from the literature are mentioned below.

Price and income. Much interest has been paid to the effect of economic incentives. Studies have generally concluded that cigarette consumption is price-sensitive, but the results with regard to cigarette quitting are more diverse. For instance, Douglas (1998) and Kidd and Hopkins (2004) report no effect of cigarette prices on quitting, whereas Forster and Jones (2001), Nicolás (2002), and Taurus (2004) find some evidence that higher prices lead to larger quitting rates. Tauras and Chaloupka (2001) find the quitting rate of young females to be slightly more price-sensitive than that of young males and Madden (2007) suggests that females with low education respond more to price changes than do females with higher education. DeCicca and colleagues (2008) report that light smokers (daily consumption 1-10 cigarettes) respond more strongly than heavier smokers, Gallet and List (2003) suggest that young smokers are more responsive to price changes than their older counterparts. More recent results, however, propose that also older adults' quitting behaviour is quite price-responsive (DeCicca and McLead 2008). Douglas (1998) and Hyland et al. (2004) find a positive income effect on cigarette cessation even after controlling for education and a range of other covariates.

Health status and health shock variables. In addition to factors like personal characteristics, addiction, prices, and income, the smoker's health condition and changes to health status could possibly affect the quitting decision. While findings in Blaylock and Blisard (1992) suggest that self-assessed health status is not important for cessation, findings in Jones (1994), Farnworth (2006) and Clarke and Etilé (2002, 2006) suggest the opposite. In a longitudinal study Hsieh (1998) did find that reduced health, defined as declines in self-assessed health and increases in the number of diseases and limitations, increased the probability of quitting. Smith et al. (2001) analyze how exogenous health shocks affects people's longevity expectations and report that smokers respond more to smoke-related health shocks than non-smokers and they were more influenced by smoke-related health shocks than by other types of health shocks. Clarke and Etilé (2002) find that smokers with worsened health and those with a heart check-up were more likely to quit. The authors interpret the findings as impersonal health information having less of an effect on smoking than personalized health information.

The rest of the paper proceeds as follows: Section 2 presents the econometric model while the data descriptions and definitions of the variables, etc., are found in Section 3. Estimation results and sensitivity analyses are presented in Section 4. Section 5 offers concluding remarks and a discussion of possible policy implications of the findings.

2. Model and method

Our sample can be described as generated by stock sampling from an underlying population (see Lancaster 1990, p. 91 and Verbeek 2004, p. 247), as the individuals are sampled conditional on being smokers at the start of the observation period. In modelling the cigarette quitting hazard, we keep attention on two time variables: time in process as a smoker and time in the screening process. The motivation is that a respondent's inclination to cease smoking is likely to depend on: how long he/she has been addicted to smoking and how long he/she has been scrutinized by health authorities and has accumulated health screening information. Time thus enters our model in a more complex way than in analyzing, say, age and time-specific mortality rates and similar demographic analogues.

We employ a *discrete-time hazard model* to examine the effects of certain covariates on the probability to quit smoking. We let t denote the running calendar time in years, index the individual smokers by i and analyse the stock of persons conditional on already being a smoker. The model now to be described resembles models for duration analysis in the literature, but since its interpretation departs from the standard setup in certain important respects, we elaborate the description in some detail.

The sample contains observations from n individuals, indexed by $i=1, 2,\dots, n$. We know that all individuals are smokers when the observation period starts, that some quit smoking at a known date during the period, and that the remaining individuals are smokers beyond the observation period. It is, however, unknown to us if and when smoke cessation then will eventually occur. The observation period extends from period $t=\tau$ till period $t=\tau+S$, the width of the observation window being S. Let s_i $(\leq S)$ denote the number of observations from individual i is at risk of quitting smoking, so that $t=\tau, \tau+1,...,\tau+s_i$ indexes his/her (observed) time series. The dataset is organised so that each person contributes the number of time units equal to the width of the observation window if he/she is still a smoker at the end of the study period (censored cases, $\delta_i=0$) or equal to the time units when they quit smoking ($\tau+s_i\leq S$, uncensored cases, $\delta_i=1$).

Further, let the stochastic integer variables B_i and T_i ($B_i < \tau < T_i$) represent the calendar periods in which individual i begins and ends smoking, respectively. The value of B_i can be recovered for all individuals in the current data set. It follows that T_i is known and equal to $\tau+s_i$ only if $s_i \leq$ S ($T_i \leq \tau$+S) and individual *i* is right-censored by the sample design if $T_i >$ τ+S*)*. We assume that non-smoking is an absorbing state: no individual having ceased smoking restarts later on.

The (discrete- time) hazard rate $h_{i,t}$ is the probability that smoker *i* quits in year *t*, conditional on (i) having started in period B_i and (ii) having smoked up to period *t*:

$$(1) \quad h_{i,t} = P(T_i = t | B_i, T_i > t-1), \qquad t = \tau, \ldots, \tau + si \ .$$

where B_i explains smoke cessation in period *t*. The corresponding probability that smoker *i* does not exit in period *t* is (1- $h_{i,t}$), and the conditional probability of observing the event history of a person who continues to smoke throughout the observation window [τ, τ+s_i] is:

$$(2) \quad q_{i,t} = P(T_i > t | B_i, T_i > \tau - 1) = \prod_{\theta=\tau}^{t} (1 - h_{i,\theta}), \qquad , \qquad t = \tau, \ldots, \tau + si \ .$$

Equations (1) and (2) imply that the probability of individual i quitting smoking during the study interval is

$$(3) \quad P(T_i = \tau + s_i | B_i, T_i > \tau - 1) = h_{i,\tau+s_i} \prod_{t=\tau}^{(\tau+s_i)-1} (1 - h_{i,t}) = h_{i,\tau+s_i} q_{i,\tau+s_i-1} \ .$$

Individual *i* ‘s part of the likelihood function can be written as $L_i = h_{i,\tau+s_i} q_{i,\tau+s_i-1}$ for $\delta_i=1$ and as $L_i = q_{i,\tau+s_i}$ for $\delta_i=0$, giving the full log-likelihood function (see, e.g., Kalbfleisch and Prentice [1980. sections 3.2, 5.2])

$$\log(L) = \sum_{i=1}^{n} \log(L_i) = \sum_{i=1}^{n} [\delta_i \log(h_{i,\tau+s_i} q_{i,\tau+s_i-1}) + (1-\delta_i) \log(q_{i,\tau+s_i})]$$

$$= \sum_{i=1}^{n} [\delta_i \log\left(\frac{h q_{i,\tau+s_i-1}}{q_{i,\tau+s_i}}\right) + \log(q_{i,\tau+s_i})]$$

Using (2) and (3), the latter can be re-expressed in terms of the hazard rate function as

$$(4) \quad \log(L) = \sum_{i=1}^{n} \{ \delta_i \log[\frac{h_{i,\tau+s_i}}{(1 - h_{i,\tau+s_i})}] + \sum_{t=\tau}^{\tau+s_i} log(1 - h_{i,t}) \} ,$$

For computational convenience we define a double indexed variable y_{it}=1 if t=τ+s_i and δ=1, and y_{it}=0 otherwise, which imply that the log-likelihood function can be rewritten as a double sum (see Jenkins 1995, section II):

$$(5) \quad \log(L) = \sum_{i=1}^{n} \sum_{t=\tau}^{\tau+s_i} \{y_{i,t} \log[\frac{h_{i,t}}{(1-h_{i,t})}] + \log(1-h_{it})\}$$

Two common choices for specification of the complementary, continue smoking probability, $1-h_{it}$, are those implied by the type I extreme value distribution and the logistic distribution. We choose the former, i.e.,

$1-h_{it} = \exp[-\exp\{-z_{it}\}] \quad \le=> \ -\log[-\log(1-h_{it})] = z_{it}$,

where z_{it} is a variable expressed as a linear function of reported and registered covariates and time in the observation period, all represented by the vector x_{it}.

The hazard of quitting may also be influenced by the number of years as a smoker (length of the smoking career) prior to entering the sample. The way we let smoke cessation depend on this is, as remarked, to allow for stepwise changes in the coefficient vector of x_{it}, with the length of the smoking career (SC) before the observation start period, τ, *i.e.* SC_i .= τ - B_i . 'Duration dependence' of smoking is thus captured through the different impact of the demographic and the health-shock variables according to the respondents' smoking career in the pre-sample period and through a set of time dummies for the observation period.

In addition to the x_{it} vector, z_{it} includes $\varepsilon_{i,}$, a random variable representing unobserved individual factors in the inclination to cease smoking. Heterogeneity in the survival processes, in the present case 'survival' as a smoker, is often denoted as 'frailty' in bio-statistics; see e.g., Hanagal, 2008). Letting β_j be the coefficient vector if the length of the pre-sample smoke-career belongs to the *j'th* interval I_j *(ST-, MT- and LT-smokers, respectively)*, and assuming ε_i, distributed independently of x_{it}, we end up with the following parameterization:

$$(6) \quad z_{it} = -\log[-\log(1-h_{it})] = \beta_j' x_{it} + e_i \quad for \ SC_i \in I_j$$

Latent heterogeneity in the cigarette-quitting hazard h_{it} is essentially different from latent heterogeneity in the non-smoking probability, 1- q_{it}. Several ways of modelling heterogeneity in survival processes have been proposed, based, inter alia, on the normal (Gaussian) distribution, the gamma distribution, and the discrete multinomial distribution. In the latter case, the mass points and the corresponding probabilities of the discrete distribution can be estimated together with the β_j 's.

3. Data

Data sources and sample construction

The main body of data for this study is extracted from a comprehensive cardiovascular screening program that was initiated in 1974 ("The Norwegian Counties Study"). The former National Health Screening Service (now Norwegian Institute of Public Health) conducted three cardiovascular screenings in three chosen counties over the 1974-1988 period[1]. In the first screening *all* inhabitants aged 35-49 years – age being dated at the first screening - and a 10% random sample of persons between 20 and 34 years old were invited. The target groups for the second and third screening were a combination of previous participants and new cohorts. The three screening dates will be denoted as R1, R2, and R3.

Altogether 65.624 subjects were invited to the first screening in the three counties, and 88% participated (Bjartveit et al. 1979). The total attendance rates to the second and third screening were 88% and 84%, respectively. All invited persons were asked to fill in a questionnaire at home and bring it to the screening station. The questions were concerned with the history of cardiovascular disease, diabetes, use of anti hypertensiva, symptoms of cardiovascular disease, physical activity during leisure time and at work, smoking habits, stress factors in social life, and family history of coronary heart disease. An additional questionnaire was handed out at the screening station and the participants were asked to fill it in at home afterwards and return it by mail. A simple health examination was carried out at the screening station. Height and weight, systolic and diastolic blood pressure were measured according to a standard protocol and a non-fasting blood sample was drawn and analysed for serum total cholesterol and triglycerides. At R1 and R2 a mass miniature chest x-ray was

[1] The counties and the screening periods were: "Oppland" 1976-1978, 1981-1983, 1986-1988; "Sogn og Fjordane" 1975-1976, 1980-1981, 1985-1986, and "Finnmark" 1974-1975, 1977-1978, 1987-1988.

taken. Results from the medical tests are used together with responses from the questionnaires.

We employ data for individuals who were screened for the first time in 1975-1978 and participated in both follow-ups. Technically, as participants provide smoking information that encompasses their smoking status also one year prior to the screening, we start the panel at 1974 for those screened in 1975, at 1975 for those screened in 1976, etc. As remarked, only respondents who stated they were daily cigarette smokers at the start of the study are included. The number of smokers participating in the three screenings is 12.499. To account for the variation in smoking careers prior to R1, however, we split the sample into three groups, based on the number of years as a smoker: 1) *Short-term smokers* (ST-smokers) have smoked up to 5 years at R1, n=905; 2) *Medium-term smokers* (MT-smokers) have smoked 12-16 years at R1, n=2,644; 3) *Long-term smokers* (LT-smokers) have smoked 25 years or more at R1, n=2034 (final sample size: n=5,583) . The average number of observations per individual is 9.82, and the three groups have 7,628; 25,364 and 20,532 observations, respectively.

For the 1974-1988 period, data from the screening is merged with annual information from administrative registers (Statistics Norway) on the individuals' income, education, marital status, and size of the household. Annual cigarette prices are obtained from the same source. The five categories of explanatory variables are defined and presented in Tables 1 and 2, with descriptive statistics in Table 3. We make extensive use of dummy variables and also let some of them interact with time, see Table 2.

The model description in Section 2, assuming that a contiguous unbalanced panel data set with s_i observations of individual i exists, is somewhat simplistic as the original sample is not strictly contiguous with respect to all variables. Rather, it is a contiguous**,** unbalanced panel data set recording more variables in the screening years than in the non-screening years, so that we actually have a mixture of a 3-wave and a 14-year panel. As soon as a person ceases smoking, he/she will be dropped from further follow-ups, even if a restart should occur later. The sensitivity of this assumption will be discussed at the end of Section 4.

Sample description

Demographic variables: Table 3 shows that there are roughly equal proportions of men and women in the total sample (dummy *male*) while large differences are found across the three subgroups. The proportion of males is relatively low among ST- and MT-smokers (28 and 44%, respectively), whereas 78% of LT-smokers are men**.** Mean age at the start of the survey is 39 years (*age*) and increases with the length of the smoking career (from 32 to 44 years across subgroups). Regarding education, more than half of the sample had left school after the mandatory minimum schooling (≤ 9 years), roughly 40% reporting secondary school as their highest level of education, while less than 1% had a university degree (dummies *educ1-educ4*) – according to information from Statistics Norway. More people (76%) had children under the age of 16 at the start of the study than later, the share being 58% for the whole period (dummy *children*). Medium-term smokers had the highest and Short-term smokers the next highest frequency of children. Among ST-smokers almost 20% were unmarried (dummy *single*), while the overall share is 10%. Table 3 shows that 6% of the sample has been separated, divorced or widowed (dummy *div-wid*) over the study period; 85% is *married.*

Indicators of addiction: As the hazard of quitting is likely to be influenced by the strength of cigarette addiction, two tobacco addiction indicators are included. The first relates to the age of onset of smoking (*debut_age*). It varies around a mean of 21 years, and declines significantly with the number of years as a smoker (29 years for ST- and 18 years for LT-smokers). The second addiction variable is the log of the maximum number of cigarettes smoked per day (*number_cig*). Again, the mean varies with smoking experience; ST-smokers report to smoke 10 cigarettes per day, MT-smokers 13 cigarettes, and LT-smokers 15 cigarettes. As an indicator of an addictive personality, and to examine a possible relationship between nicotine and caffeine, we also include a dummy indicating whether or not the individual drinks more than 9 cups of coffee per day (*much_coff*).

Price and income indicators: The cigarette prices and income variables, also potentially important factors motivating smoke cessation, are defined as follows: First, we let the annual log-increase of the CPI-deflated price of a 20-pack of Marlboro represent the price development for cigarettes in general ($year_t$ - $year_{t-1}$), as prices of the various cigarette brands in Norway tend to move almost in parallel. Our motivation for representing price variables in this way is that since current smokers can be said to have already “absorbed” the previous year’s price level, it is the relative price increase that has a potential impact on the quitting hazard. Second, considering the high proportion of married respondents and female home-makers with unpaid employment, family income is preferred to individual income [2].

Calendar variables: Time variables enter the model in several ways. First, we include a full set of time dummies for the years in the observation period to represent individuals' `being in the smoking process', i.e., the `baseline hazard' in duration analysis. Second, to account for the length of the pre-sample smoking career we have, as mentioned, divided the sample into three groups according to the information given on the number of years as a smoker at the first screening. In addition are included dummies for the three screening years, which, unlike the time dummies, are *(i,t)*-subscripted, because the screening period is not synchronized for the sampled individuals. They account for `being in the screening process', confer the introduction. Other potentially important time variables for smoking cessation, like age of smoking initiation and cumulated addictive stocks, and cumulated (information on) health stocks are represented, more indirectly, via addiction and health-shock variables.

Health status and health-shock indicators. General remarks: The way the respondents' information about own health status affects the smoking cessation is one core issue of this study. Our primary motivation for the chosen variable specification is based on our assumption of how the sequence of events as emerging to the respondents - his or her `information set' becoming gradually larger – affects behaviour. It can, briefly, be described as follows: At the start of the study period the respondents have a certain health status and a stock of information about it. At R1, R2, and R3 some new information is provided, which is supplemented by results from the medical tests announced afterwards. This information - which may to (at least) some respondents emerge as health shocks - may affect the decision to continue as a smoker or to quit. We have incorporated these ideas by including: (i) health status variables at R1 (self-reported and registered by health personnel), (ii) dummies indicating worsened health status from R1 to R2, or from R2 to R3; (iii) test results from blood tests and x-rays interacting with dummies for the corresponding screening year only.

[2] Statistics Norway has provided us with the relevant family income data (based on assessments for tax purposes). We have deflated income by the CPI and normalized it by family size by division with the square root of the number of persons, including the children below 16 years of age. For a small subsample, income data was missing or reported as zero in certain years. Instead of deleting these units, at the risk of introducing selection bias, we have replaced the missing values by the individual's mean income in the other years of the study period.

Health status indicators: Symptoms of a cardiovascular illness at R1 (indicated by the dummy *symptm*) were reported by 2.6% of the respondents, 4% reported to have/have had a cardiovascular disease or diabetes at R1 (dummy *illness*), and one third had a body mass index (BMI) above 25 at that time (dummy *bmihigh*). A significantly higher proportion of LT-smokers than of smokers in the other groups had experienced such illnesses and had a high BMI. Long-term smokers also reported a higher frequency of a physically demanding job (dummy *actwork*): 47% versus 31% for ST-smokers, and a higher frequency of exercise (dummy *exercise*) than the group having a smoking career less than 6 years. Around 2% of the sample received disabled pension (dummy *disabled*).

Health shock indicators: Dummy variables for negative "health changes" constitute the second group of health indicators. Obviously, changes in symptoms or illnesses recorded at R2 and R3 may have started to influence smoking behaviour prior to that date (the individuals may e.g., have experienced symptoms indicating lung problems or contracted a cardiovascular disease soon after the previous screening). As described in Table 2, this is taken into account when creating the *sympchange12-23* and *illchange12-23* variables. Among those who had not reported any cardiovascular diseases or diabetes at R1, 12% did so at R2 or R3 and symptoms for the diseases were recorded for an additional 4% during the study period. Long-term smokers reported a higher prevalence of symptoms or illnesses at R2 and R3 than the other two groups. If the results from the blood samples, measurement of blood pressure, or x-rays came out with a score over a critical cut-off point (set by medical experts), the screened person was contacted with the recommendation of a follow-up. We assume that if a negative test result would affect the individuals' smoking behaviour, it would do so the same year as the news was received. Thus, we have created dummy variables that are equal to one, for the respective screening years only, if a negative test result is revealed to the participants (*badreport1-3*). As the possible effect of the test results on smoking participation only will be registered at the next screening, the *badreport3* variable is left out of the analyses.

Intervention variables: Our last set of explanatory variables is dummies for the screening years (*screening1-3*). After controlling for all the variables described above, we want to explore whether there is a remaining effect that can be attributed to the screening itself. The idea is that the smokers, through the attention raised by the screening invitation, participation, and the testing (and irrespective of the test results), are reminded of and confronted with their individual health risks and that this increased attention may cause changes in smoking behaviour. The screenings can be viewed as a form of governmental intervention that comes in addition to the excise tax on tobacco accounted for in the price variable.

Dependent variable: The dependent variable in the duration analyses is the time-varying dummy for quitting. Over the study period 430 individuals (48%) among the group of ST-smokers ceased smoking, while the corresponding numbers for MT- and LT- smokers were 898 (34%) and 601 (30%), respectively. The quitters were asked whether they had terminated the habit less than 3 months, 3-12 months, 1-5 years or more than 5 years prior to date of recording. From these entries we have constructed a binary quitting variable based on the year the individuals ceased smoking (if they quitted smoking at all). For the 1-5 years category the quitting year is constructed from the information the respondents have given regarding the number of years as a smoker at each screening[3].

4. Results

Before presenting estimation and test results, we consider the non-parametric hazard rates for ST-, MT- and LT-smokers in Figure 2. Its x-axis represents the time elapsed since the start of the study. Because this start is not synchronized for the respondents, the first period represents one year between 1974 and 1977. Despite varying hazard levels, all figures exhibit a downward trend and a three-peaks pattern. The three peaks may reflect the three rounds of screening although there are possible alternative explanations. For all three groups of smokers, period 2 represents R1 (the year of the first screening), R2 occurred in period 7 for the majority of the respondents, but varies somewhat, as does the period when the screening was terminated, R3.

A closer examination of the "peaks" reveals that among the 631 smokers who quitted in the first screening year, 41% reported to have done so within the 3 months period before visiting

[3] For a small fraction of the respondents (1.9%) this information could not be used due to obvious measurement errors (e.g., reported fewer years as a smoker at R2 than at R1). An additional 5.9% reported a number of years as a smoker which suggested a "heaping effect" (reported 5, 10, 15,. .. years at R1, R2 and R3) that gave the value "0" when subtracting years at R2 from years at R1, when the interval between the two screenings were 5 years. To avoid possible selection bias, these subjects (8.8%) are assigned a randomly picked year of smoking cessation within the 1-5 year interval. To test the sensitivity of the assignment we re-ran the estimations for the three groups excluding this subgroup of quitters. The estimates remained roughly the same, with the identical demographic and addiction variables being statistically significant. The hazard ratio for the income variable for LT-smokers changed from 0.812 to 0.862 but was no longer significant, neither was the Symptoms at R1 variable. For MT-smokers a high BMI and exercising at R1now became significant at a 10% level. The hazard ratio of screening remained highly significant for all groups and its value increased.

the screening station, 24% quitted 3 to 12 months before the screening while 35% quitted smoking *after* the screening that year. The corresponding numbers for the second screening year are 36%, 30%, and 34%, respectively. As the third screening year marks the end of the follow-up study, only those who quitted before visiting the screening station were registered as quitters. Of the 217 people claiming to have ceased smoking that year, 148 reported to have done so within the last 3 months, whereas 69 had quitted earlier in the year.

The following analysis will reveal to what extent this distinct pattern of the hazard can be accounted for primarily by screening participation, or whether other factors, e.g., price changes, can explain the major part of its variation.

Model selection

Estimation is performed by maximizing the log-likelihood function, as given by (5), employing routines in the *Stata11* software. We use the cloglog module for standard binary choice, as well as its extension allowing for unobserved individual specific heterogeneity (frailty), exploiting the unbalanced design of the panel data set. For the pooled sample three distributions of unobserved heterogeneity - normal, gamma, and discrete - were considered. Likelihood Ratio (LR) tests, reported in Table A in Appendix, indicate that the model performs better when unobserved heterogeneity is accounted for. The Akaike's information criterion (AIC) and the Bayesian information criterion (BIC), also reported in the Appendix, suggest that the gamma specification is superior. Therefore, the cloglog model with gamma distributed heterogeneity will form the basis for the results presented below.

Estimation results - overall

The following discussion is organized by categories of covariates, and for expository convenience, the hazard ratio estimates and other statistics are split into two tables, Table 4 and Table 5. The influence of personal characteristics, indicators of addiction status, and economic factors are presented in the former. The latter contains results for health status and health shock variables and indicators of government interventions, as well as goodness of fit statistics and statistics describing the latent heterogeneity. The entry "Exp(b)" in the table heads indicates that the results are the anti-logs of the coefficient estimates, implying that the reported figures give the multiplicative effects of each specific covariate on the quit hazard rate.

Personal characteristics, indicators of addiction and economic factors

Table 4 shows that gender (dummy *male*) has a statistically significant influence on the hazard of quitting in all three "smoking-career categories". We see that the effect of gender is particularly strong among LT-smokers where the hazard of quitting for males is 5 times that of females. In comparison, the hazard ratio estimate for males is roughly 2 for the other two groups of smokers. Age seems important only for ST-smokers, and we find that younger people have a higher risk of quitting than have older people. As expected, we generally find that higher levels of education tend to increase the probability of quitting, although for LT-smokers, only the highest level of education comes out as statistically significant.

The coefficient for having children younger than 16 years is small and statistically insignificant for all the three groups of smokers (dummy *children*). Table 4 further suggests that among LT-smokers, married people have a higher quitting hazard than single, divorced, and widowed individuals (the hazard ratio for the dummies *single* and *Div-wid* have a value less than one). Also for ST- and MT-smokers, is being single associated with a lower quitting probability. Being widowed or divorced, on the other hand, does not seem to influence the quitting behaviour in these groups as the estimates here are not statistically significant.

Three of the covariates included are intended to represent addiction, two relating to nicotine. Also here we find pronounced differences between the smoking career categories. For ST-smokers both of these indicators come out as statistically significant: the older a person is when starting to smoke, the more likely it is that he/she will quit later on (*debut_age*), and the higher number of daily cigarettes consumed (*no.cig*) the less is the quitting hazard. For MT-smokers only the latter is significant, whereas for LT-smokers none of the two tobacco addiction indicators are. It has been hypothesized that addiction to caffeine and nicotine interfere - the two being either complements or substitutes. The negative coefficient of the 'much coffee' dummy – although statistically significant at a 10% level only for LT-smokers – supports the complementarity hypothesis: the same people tend to be addicted to both. Drinking more than 9 cups of coffee per day could suggest an "addictive personality".

Changes in the deflated cigarette price do not seem to influence the quitting decision neither for ST nor for LT-smokers. The statistically weak estimate for MT-smokers indicates, on the other hand, that higher cigarette prices increase the probability of smoking cessation for this group. Interestingly, the income coefficient is positive and statistically significant for ST and negative and significant for LT-smoker. So even after controlling for gender, education, health, etc., higher family income increases the likelihood of quitting for Short-term smokers and decreases it for Long-term smokers. For MT-smokers, the income effect is very small in absolute value and statistically insignificant.

Health status variables, health shock indicators and governmental interventions

Maybe the most novel results in this study are those reported in Table 5. Overall, one striking feature is the increasing strength of response of the quitting hazard to the health status variables with increasing smoking experience; compare the blocks of hazard ratio estimates in the table. The quitting hazard of ST-smokers, who are unlikely to have yet experienced severe negative health consequences of their habit, is not significantly influenced by any of the health indicators reported. Neither reported symptoms, nor illness, nor a high body mass index (BMI) at R1, nor hard physical work, regular exercise nor receiving disabled pension seem to affect the quitting probability for this group. The same is the case for MT-smokers, whereas for LT-smokers the hazard of quitting increases substantially when cardiovascular symptoms and a high BMI are reported at R1. Having a physically demanding job and receiving disabled pension reduce the likelihood of ceasing smoking for people who have smoked for more than 25 years.

Turning to the health shock variables, we find that when the study participants report symptoms of heart or lung problems between two adjacent screenings (dummy *sympshock12, sympshock23*), the quitting hazard shows no statistically significant response for any of the smoking groups. On the other hand, if for MT- and LT-smokers an illness has occurred between the screenings (dummies *illshock12, illshock23*), he/she tends to respond by significantly increasing the cessation rate. For MT-smokers, the effect seems particularly strong, while for LT-smokers only one of the two latter health shock indicators is statistically significant. Receiving a bad result on any of the blood tests or the x-ray taken at R1 or R2 does not seem to influence the cessation rate (*badreport1 and 2*).

A finding of particular interest is that the coefficients for the intervention variables (dummies *screening1,2,3*) in Table 5 are positive and highly significant in most cases. This suggests that the peaked hazard rates pattern in Figure 2 is not spurious: the impact of the screenings emerges also when conditioning on all the covariates accounted for above. The coefficient of *screening3* for LT-smokers, however, is not statistically significant. As the study is designed such that the dating of this variable coincides with the end of the observation period, any influence on the hazard must come only from people ceasing smoking prior to the screening (not including those who quit shortly after they were screened), which may partly explain the lack of influence of this intervention dummy.

Latent heterogeneity

Latent heterogeneity, supplementing observed heterogeneity represented by cohort, gender, marital status, addiction status, etc., comes out as significant and important. For every group of smokers and for all the parametrizations of the heterogeneity considered, the null hypothesis of no random heterogeneity is rejected (Table 5). For MT-smokers, however, the LR-test statistic is significant only at an 8% rejection level, whereas the effect for LT-smokers seems particularly high. Consequences of erroneously ignoring unobserved heterogeneity are particularly important because the model is non-linear. This subject, with respect to health economics applications of the hazard rate models, is discussed in Jones et al. (2007, p. 193).

Sensitivity analyses

Definition of "true quitters": Since many smokers who give up smoking are known to re-start again at some later point, the results above, from a sample also including quitters for which a relapse occurs later on, are potentially biased. However, although the individual time-series in our sample are "cut off" when the respondents report to have ceased smoking, the panel structure of the data allows us to examine to what extent a relapse occurs. Doing this, we found that 27% of the smokers who quitted in the first screening year reported to be daily cigarette smokers again at R2 and 25% of quitters in the R1-R2 period had started again at R3. To assess the magnitude of the potential estimation bias, we re-ran the models after having excluded *all* observations from the quitters for whom a relapse to smoking is known to have occurred at some later point. Relative to the results reported in Tables 4 and 5, the impact of the first screening program was then somewhat diminished for all groups (e.g., the hazard ratio was reduced from 3.706 to 2.428 for ST, Table 5), the impact of the second screening increased for MT- and LT-smokers and became statistically insignificant for ST ($p<0.17$), while the effect of the third screening was virtually unchanged for MT- and LT-smokers and reduced for ST-smokers. However, even though the estimated influence of screening was reduced in some cases (somewhat lower coefficients or higher p-values), we find it still safe to conclude that being invited to and participating in the program is important for the overall impetus to quit among all groups of smokers in the sample.

The redefinition of the “true quitters” also had some minor consequences for the remaining covariates (confer Table 4), with one exception: having a physically demanding job now came out as significant ($p<0.10$). Health status at the start of the study became more important for MT-smokers (more variables with a statistically significant association: illness and exercise at R1 [$p<0.05$], a high BMI [$p<0.10$]), while cigarette prices became insignificant. The coffee indicator and being widowed/divorced were no longer statistically significant for the quitting behaviour among LT-smokers. Overall, the results seemed fairly robust.

Influence of the screening dummies: Given the seeming importance of the screening dummies on the quitting hazards we also wanted to examine to what extent excluding them from the model would influence the estimated effect of the remaining covariates, by re-running the estimations without these dummies. The results were very similar to those in Tables 4 and 5; the same covariates being statistically significant and of basically the same magnitude. The coefficients of the time dummies change somewhat more, in particular the dummy for period 7 (corresponds to R2 for many respondents) switched from being small and non-significant to becoming higher and significant.

The overall conclusion of this sensitivity analysis is that the main results are fairly robust to alternative model specifications and definitions of quitters.

5. Concluding remarks and policy implications

This paper presents a (discrete-time) duration analysis of factors motivating daily cigarette smokers to quit, and contributes to the literature in mainly three ways. First, it provides new estimates of the effects of a wide range of covariates and traditional policy means. Second, it offers evidence on alternative interventions for health authorities to increase the cessation rate among daily cigarette smokers. Third, its specific focus is on examining and comparing three groups of smokers, including one that has been smoking for more than 25 years at the time of the first screening. The duration model applied has a cloglog form and allows for flexibility both in the representation of time – including the time in process as a smoker and time in the screening process – and in the way unobserved individual heterogeneity is accounted for.

The results clearly confirm that the majority of the covariates have effects, which differ according to the length of the pre-sample smoking career, but also that total time in process is important. The influence of personal characteristics and addiction indicators on the quitting hazard is highest among ST-smokers, MT- and LT-smokers being less strongly affected. The longer the duration of the pre-sample smoking-career, the smaller impact do the age at initiation and the number of daily cigarettes seem to have on quitting, suggesting that a long-standing habit may be equally difficult to break irrespective of the smoking intensity. On the other hand, our finding that the smoking initiation age significantly affects the quit hazard of ST-smokers, suggests that anti-smoking policies succeeding in delaying smoking onset, may increase the quitting hazard at some later point in time for this group. For ST- and MT-smokers, policies that contribute to a less intensive smoking habit (lowering the number of daily cigarettes smoked) may pull in the same direction.

Adverse health outcomes recorded at the start of the intervention period appear to influence the cessation rate only for LT-smokers. Neither having a cardiovascular disease or diabetes – or symptoms of such – nor being disabled, having a high BMI, a physically demanding job or exercising regularly are associated with an increased quitting hazard for ST- or MT-smokers. No group of smokers seems to significantly increase their quitting hazard in response to a bad test result, which is in line with findings in a study by van der Aalst and colleagues (2010). On the other hand, health shocks recorded during the intervention period significantly increase the quitting hazard for MT- and LT-smokers, the hazard ratio being higher for the former than for the latter. This may suggest that a smoker is more responsive to health information the longer he/she has been smoking.

While it is generally assumed that young smokers are more price-sensitive (Farrelly et al. 2001), it is still a matter of dispute whether adult quitting behaviour is influenced by price increases (DeCicca and McLeod 2008). Representing cigarette prices by a relative price increase variable, we find that only MT-smokers are responsive to price changes, and only at a 10% level. However, even if the cigarette tax instrument does not seem to have a strong effect on smoking cessation, its assumed effect on smoking intensity (Chaloupka and Warner 2000; Gallet and List 2003) may indirectly increase the quitting hazard rate. We find that income significantly influences the smoking cessation hazard of ST- and LT-smokers, but in opposite direction: an increase for the former, a decrease for the latter. The latter result for LT-smokers is not easy to explain. Furthermore, education had very little impact on quitting for this group, which may also seem surprising.

The strong and significant effect on the quitting hazard of the interventions represented by the screening dummies, considering the other ways time is represented in the analysis, is interesting and surprising. As shown above, many of those who ceased cigarette smoking in the screening years reported to have done so during the three months immediately preceding the participation date. One interpretation of this finding is that the letter of invitation reminded and alerted the smokers of the negative health effects of their cigarette consumption and raised a fear for what the screening could possibly reveal. So for a fraction of the smokers, having perhaps already considered giving up the habit, the reminder seems to have been sufficient to actually take action. Also, the quitting hazard shortly after participation was high, which is in line with a meta-analysis of the efficacy of tobacco counselling by health care professionals group, which showed that a significant increase in cigarette quitting occurred after physicians, nurses, and others had given patients cessation advice (Gorin and Heck 2004). The finding of a screening effect is also in line with results reported in a small study of the effect of a CT screening on smoking cessation (Ostroff et al 2001).

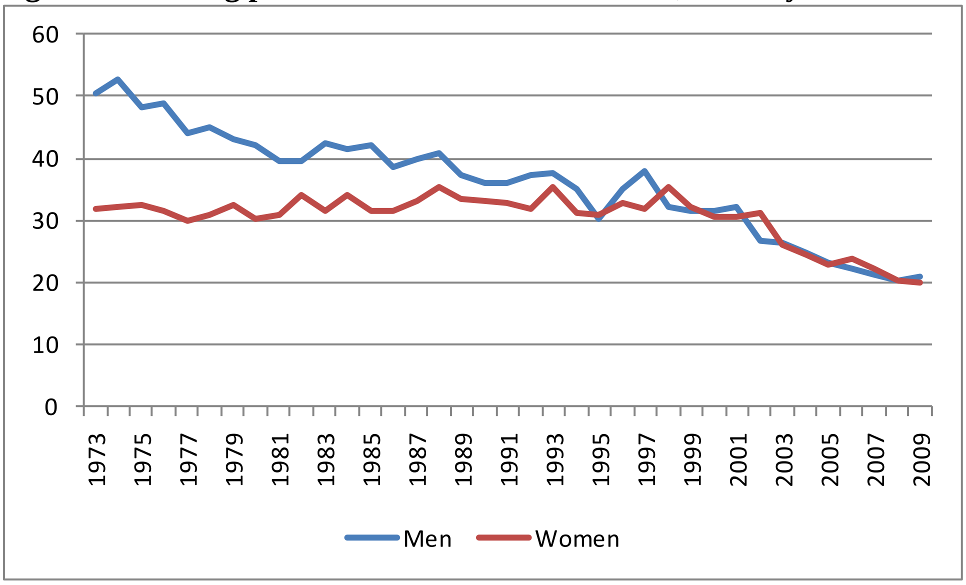
Since by general belief, people having smoked for more than 25 years are more “immune” to anti-smoking interventions than smokers with a shorter career, our coefficient estimates for the LT-smokers are worth noting, in particular to policy makers and others trying to promote improved health in the population. Increasing the quitting hazard in this group of smokers is definitely important since the health gains for giving up smoking are substantial even for smokers with a long-standing career (Taylor et al. 2002; Ostbye and Taylor 2004). Thus, an extended use of targeted screening programs, tobacco counselling by health care providers, etc., seems especially useful for this group of experienced smokers.

Our data set, although being far from perfect, has a rather long observation window, many participants, combines personal characteristics, indicators of addiction status, economic factors, health status and health shock variables, and governmental interventions within the same data file. This may suggest that problems related to omitted variables bias, spurious effects interactions, etc., are less pronounced than in similar studies based on shorter data vectors. Also, the fact that data from administrative registers lie `at the bottom' may reduce certain measurement problems (recall bias, etc.). However, being a panel data set, its main limitation may be that its records are from only a three-wave panel within a fourteen-year period, by necessity giving less information about the two intervening periods. Further, as some smokers restart smoking after periods of cessation, the estimated effects may be considered as upper bounds. Anyway, the significant effect of being invited to and participating in a screening as well as the other findings seem robust, which may prove useful in assessing future policies to promote smoking cessation.

Acknowledgement

We would like to thank Karl Erik Lund, SIRUS, for useful comments and suggestions on a previous version of the paper.

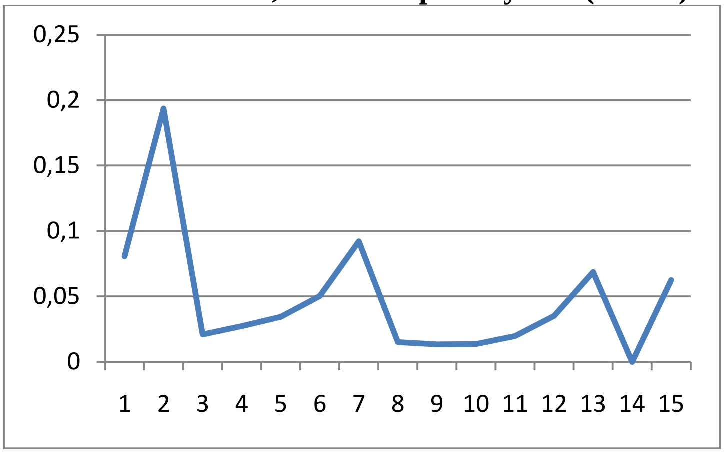
Figure 1 Smoking prevalence for men and women, Norway 1973-2009



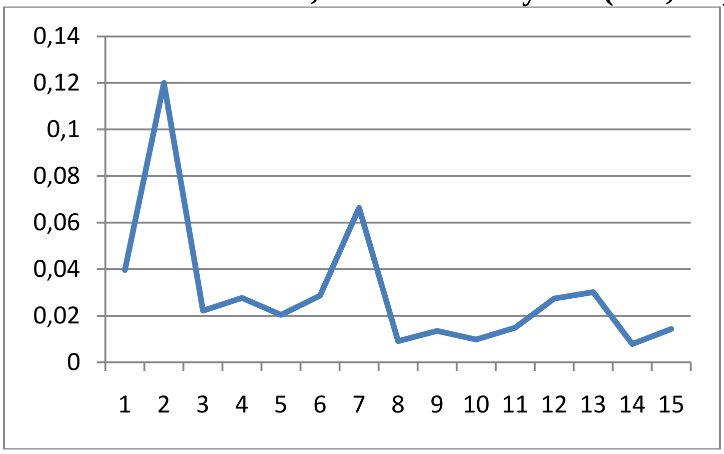
Source: http://www.ssb.no/royk/

Figure 2 Hazard rates of cigarette quitting*

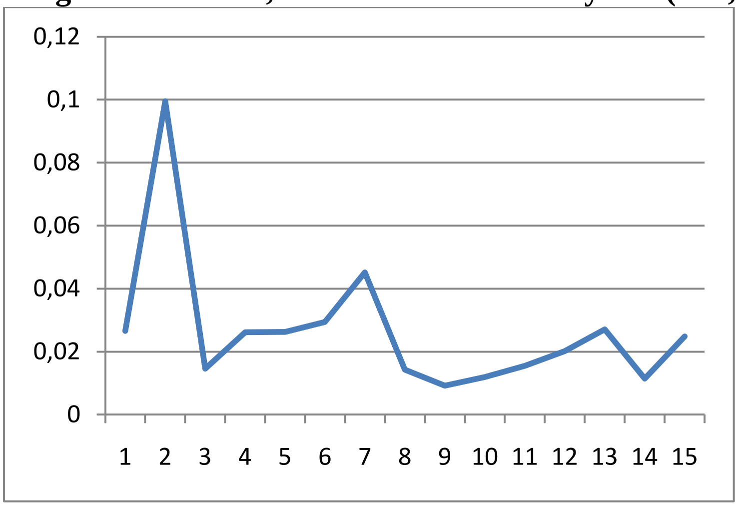
Short term smokers, smoked up to 5 years (n=905)



Medium term smokers, smoked 12-16 years (n=2,644)



Long term smokers, smoked more than 24 years (n=2,034)



***The x-axis represents the number of time periods (years) since the start of the study

Table 1. Variable description

Variables	Operationalisation	Type*	Data source**
Demographics			
Male	Dummy;1 if male	2	H.P, screening
Age	Age at start of survey 1	Time invariant	S.R, screening
Age*male	Interaction of age at R1 and male	Time invariant	screening
Educ1	Dummy; 1 if highest education is min. schooling (mandatory)	3	Statistics Norway
Educ2	Dummy; 1 if highest education is secondary school	3	Statistics Norway
Educ3	Dummy; 1 if between 12 and 15 years of schooling	3	Statistics Norway
Educ4	Dummy; 1 if highest education is university degree	3	Statistics Norway
Single	Dummy; 1 if not registered with spouse or cohabitant	3	Statistics Norway
Married	Dummy; 1 if married	3	Statistics Norway
Div-wid	Dummy; 1 if divorced, separated or widowed	3	Statistics Norway
Children	Dummy; 1 if having children under the age of 16	3	Statistics Norway
Addiction			
Smokeage	Age when started to smoke	2	S.R, screening
Number_cig	Ln of max reported cigarettes smoked per day	Time invariant	S.R, screening
Much_coff(R1)	Dummy; 1 if drinking daily more than 9 cups of coffee (R1)	2	S.R, screening
Economic			
Cigprice	Ln of the difference of CPI adjusted price (P_{1975}-P_{1974})	Time varying	Statistics Norway
Income	Ln of CPI adjusted family income	Time varying	Statistics Norway
Health			
Symptm (R1)	Dummy; 1 if symptoms of heart/lung illness (R1)	2	S.R, screening
Illness (R1)	Dummy; 1 if having heart/lung illness (R1)	2	S.R, screening
Bmihigh (R1)	Dummy; 1 if body mass index >25 (R1)	2	H.P, screening
Actwork(R1)	Dummy; 1 if having physical demanding work (R1)	2	S.R, screening
Exercise (R1)	Dummy; 1 if exercising at least 4 hours per week (R1)	2	S.R, screening
Disabled (R1)	Dummy; 1 if receiving disability benefit (R1)	2	S.R, screening
Sympchange12	Dummy; 1 if new symptoms are reported in R2	4	S.R, screening
Sympchange23	Dummy; 1 if new symptoms are reported in R3	5	S.R, screening
Illchange12	Dummy; 1 if new heart/lung illnesses are reported in R2	4	S.R, screening
Illchange23	Dummy; 1 if new heart/lung illnesses are reported in R3	5	S.R, screening
Badreport1	Dummy; 1 if score above cut-off, blood tests or x-ray, in R1	6	H.P, screening
Badreport2	Dummy; 1 if score above cut-off, blood tests or x-ray, in R2	7	H.P, screening
Interventions			
Screening1	Dummy; 1 in the time period of R1	6	H.P, screening
Screening2	Dummy; 1 in the time period of R2	7	H.P, screening
Screening3	Dummy; 1 in the time period of R3	8	H.P, screening
Dependent var.			
Quit smoking	Dummy; 1 if quitting smoking	1	S.R, screening

*For dummy variables, see Table 2 for explanation
** S.R= self reported, H.P = registered by health personnel

Table 2 Types of dummy variables

Time period	Year	1[i]	2[ii]	3[iii]	4[iv]	5[iv]	6[v]	7[v]	8[v]
1	1974	0	1	0	0	0	0	0	0
2	1975, 1. screening	0	1	0	0	0	1	0	0
3	1976	0	1	0	1	0	0	0	0
4	1977	0	1	0	1	0	0	0	0
5	1978	0	1	0	1	0	0	0	0
6	1979	1	1	0	1	0	0	0	0
7	1980, 2. screening	.	1	0	1	0	0	1	0
8	1981	.	1	0	0	1	0	0	0
9	1982	.	1	0	0	1	0	0	0
10	1983	.	1	1	0	1	0	0	0
11	1984	.	1	1	0	1	0	0	0
12	1985	.	1	1	0	1	0	0	0
13	1986, 3. screening	.	1	1	0	1	0	0	1
14	1987	.	.	.	.	.	.	.	.
15	1988	.	.	.	.	.	.	.	.

[i]) The vector exemplifies an individual who quits smoking in 1979.
[ii]) The vector exemplifies a dummy that is time-invariant for the whole period (e.g. male, started to smoke at an early age, etc.)
[iii]) The vector exemplifies an individual who changes status in 1983(e.g. become divorced this year)
[iv]) The vectors exemplifies an individual who has changed health status between two screenings (*Sympchange12*, *23* and *Illchange12*, *23*)
[v]) The vector exemplifies an individual who has been screened in 1975, 1980 and 1986

Table 3 Summary statistics, subgroups and full sample

	Full sample		Short term smokers (Smoked up to 5 years)		Medium term smokers (smoked 12-16 years)		Long term smokers (smoked ≥25 years)	
Variable	**Mean**	**St.dev**	**Mean**	**St.dev**	**Mean**	**St.dev**	**Mean**	**St.dev**
Demographics								
Male	0.528	0.499	0.280	0.449	0.436	0.496	0.783	0.412
Age	38.91	5.815	32.35	8.904	37.38	4.275	44.45	2.813
Educ1	0.528	0.499	0.403	0.490	0.487	0.500	0.619	0.486
Educ2	0.418	0.493	0.528	0.499	0.447	0.497	0.346	0.476
Educ3	0.043	0.204	0.062	0.242	0.055	0.227	0.024	0.154
Educ4	0.006	0.075	0.001	0.032	0.007	0.085	0.005	0.069
Children	0.575	0.494	0.606	0.489	0.654	0.476	0.415	0.493
Single	0.092	0.289	0.190	0.392	0.080	0.271	0.092	0.290
Married	0.844	0.363	0.736	0.441	0.861	0.346	0.839	0.367
Div-wid	0.063	0.244	0.071	0.257	0.059	0.235	0.067	0.250
Addiction								
Debut_age	21.43	5.495	28.72	9.195	22.94	4.215	17.63	2.662
Number_cig	13.50	6.340	10.03	5.656	12.76	5.739	15.04	6.687
Much_coff(R1)	0.359	0.480	0.347	0.477	0.347	0.476	0.377	0.485
Economic								
Cigprice (ln)	1.024	5.630	0.855	5.595	1.008	5.635	1.062	5.617
Income (ln)	5.914	0.891	5.625	1.135	5.996	.8550	5.936	0.850
Health								
Sympt. (R1)	0.026	0.159	0.023	0.151	0.019	0.135	0.026	0.159
Illness (R1)	0.042	0.201	0.025	0.155	0.029	0.167	0.092	0.289
Bmihigh (R1)	0.350	0.477	0.318	0.466	0.311	0.463	0.461	0.498
Actwork(R1)	0.358	0.385	0.311	0.463	0.354	0.478	0.467	0.499
Exercise (R1)	0.203	0.402	0.159	0.366	0.202	0.401	0.202	0.402
Disabled (R1)	0.017	0.129	0.017	0.129	0.011	0.103	0.031	0.174
Sympchange12	0.008	0.092	0.011	0.104	0.006	0.080	0.013	0.115
Sympchange23	0.008	0.087	0.007	0.085	0.006	0.077	0.008	0.091
Illchange12	0.017	0.130	0.014	0.116	0.010	0.101	0.036	0.187
Illchange23	0.022	0.147	0.013	0.115	0.016	0.125	0.038	0.191
Badreport1	0.017	0.130	0.013	0.113	0.015	0.120	0.025	0.155
Badreport2	0.016	0.126	0.012	0.111	0.014	0.112	0.023	0.151
Badreport3	0.008	0.101	0.006	0.102	0.007	0.100	0.009	0.114
Interventions								
Screening1	0.098	0.297	0.109	0.311	0.099	0.299	0.096	0.294
Screening2	0.0781	0.268	0.076	0.266	0.078	0.269	0.078	0.268
Screening3	0.069	0.254	0.065	0.246	0.070	0.255	0.069	0.254
Dep. Var.								
Quit smoking	0.034	0.182	0.057	0.231	0.036	0.186	0.030	0.172
No of obs.	122,684		7,628		25,364		20,610	
No of persons	12,499		905		2,644		2,034	

Table 4 Estimation results for frailty models with gamma distribution, part 1[i].
Dependent variable is the hazard of cigarette quitting

	Short term smokers (smoked up to 5 years) No of obs.= 7,628)			**Medium term smokers** (smoked 12-16 years) No of obs.=25,364			**Long term smokers** (smoked ≥25 years) No of obs.=20,610		
	Exp(b)	**95% confidence interval**		**Exp(b)**	**95% confidence interval**		**Exp(b)**	**95% confidence interval**	
Demographics									
Male	2.036***	*1.247*	*3.325*	2.111***	*1.622*	*2.749*	5.009***	*2.518*	*9.966*
Age	0.844***	*0.745*	*0.956*	1.019	*0.960*	*1.083*	1.051	*0.976*	*1.132*
Educ2	1.565**	*1.094*	*2.239*	1.328***	*1.125*	*1.568*	1.280	*0.898*	*1.825*
Educ3	1.940*	*1.060*	*3.550*	2.004***	*1.465*	*2.743*	1.439	*0.505*	*4.102*
Educ4	12.06**	*1.409*	*103.2*	2.589***	*1.304*	*5.141*	12.11**	*1.301*	*112.7*
Children	0.864	*0.638*	*1.169*	0.968	*0.793*	*1.181*	1.097	*0.820*	*1.468*
Single	0.553***	*0.360*	*0.850*	0.698**	*0.504*	*0.968*	0.542*	*0.288*	*1.018*
Div-wid	0.629	*0.321*	*1.234*	0.900	*0.622*	*1.301*	0.538*	*0.282*	*1.026*
Addiction									
Debut_age	1.188***	*1.051*	*1.344*	1.006	*0.946*	*1.070*	1.025	*0.946*	*1.110*
No.cig	0.668**	*0.470*	*0.949*	0.639***	*0.520*	*0.786*	0.974	*0.669*	*1.417*
Much_coff(R1)	0.798	*0.571*	*1.114*	0.921	*0.783*	*1.084*	0.723*	*0.506*	*1.034*
Economic									
Cigprice	1.007	*0.985*	*1.029*	1.012*	*0.999*	*1.026*	0.995	*0.978*	*1.011*
Income	1.188**	*1.027*	*1.374*	1.003	*0.913*	*1.100*	0.812**	*0.687*	*0.959*

[i] Please note: The constant term and the coefficients of the time dummies are suppressed but available upon request. * p<0.1; **p<0.05; ***p<0.01

Table 5 Regression results for frailty model with gamma distribution, part 2[i)]
Dependent variable is the hazard of cigarette quitting

	Short term smokers (smoked up to 5 years) No of obs.= 7,628)			**Medium term smokers** (smoked 12-16 years) No of obs.=25,364			**Long term smokers** (smoked ≥25 years) No of obs.=20,610		
	Exp(b)	**95% confidence interval**		**Exp(b)**	**95% confidence interval**		**Exp(b)**	**95% confidence interval**	
Health									
Symptoms (R1)	1.380	*0.531*	*3.590*	0.991	*0.570*	*1.721*	3.202**	*1.252*	*8.185*
Illness (R1)	0.493	*0.209*	*1.768*	1.371	*0.927*	*2.027*	0.997	*0.548*	*1.812*
Bmi_high (R1)	1.294	*0.860*	*1.946*	1.131	*0.955*	*1.340*	1.850***	*1.257*	*2.722*
Actwork(R1)	0.692	*0.454*	*1.056*	1.075	*0.910*	*1.270*	0.605**	*0.402*	*0.910*
Exercise (R1)	1.151	*0.749*	*1.771*	1.160	*0.971*	*1.386*	0.975	*0.641*	*1.481*
Disabled (R1)	1.609	*0.364*	*7.122*	1.056	*0.481*	*2.316*	0.254**	*0.085*	*0.756*
Sympshock12	0.764	*0.207*	*2.815*	0.843	*0.263*	*0.843*	0.994	*0.381*	*2.593*
Sympshock23	0.899	*0.103*	*7.833*	0.591	*0.080*	*0.591*	0.700	*0.133*	*3.680*
Illshock12	1.210	*0.384*	*3.814*	1.945**	*1.068*	*1.945*	1.026	*0.604*	*1.741*
Illshock23	1.379	*0.288*	*6.612*	3.281***	*1.873*	*3.281*	1.936**	*1.039*	*3.609*
Badreport1	0.951	*0.537*	*1.687*	1.076	*0.775*	*1.494*	0.712	*0.463*	*1.096*
Badreport1	1.059	*0.495*	*2.267*	0.989	*0.629*	*1.556*	0.898	*0.529*	*1.524*
Interventions									
Survey1	3.706***	*2.315*	*5.932*	3.155***	*2.432*	*4.092*	6.125***	*3.840*	*9.768*
Survey2	3.743***	*1.527*	*9.170*	3.822***	*2.055*	*7.110*	7.634**	*3.702*	*15.74*
Survey3	3.937**	*1.203*	*12.88*	2.929***	*1.432*	*5.991*	1.090	*0.552*	*2.153*
Gamma varience	2.102**	*0.847*	*5.216*	0.537	*0.063*	*4.548*	6.260***	3.623	10.818
LR-test statistic	6.826 (P>=0.004)			2.028 (P>=0.077)			25.490 (p>0.000)		
Log likelihood	-1456.562			-3547.662			-2599.247		

[i)] Please note: The constant term and the coefficients of the time dummies are suppressed but available upon request. * p<0.1; **p<0.05; ***p<0.01

Appendix

Model selection

As noted above, quitting may be influenced by also unobserved factors. Therefore, first, we examined whether the model specification accounting for unobserved heterogeneity gave better fits than a model without such heterogeneity (Model1). We tested against a model assuming a normal distribution, a gamma distribution and a discrete distribution (Model 2-5). Next, as the model not accounting for unobserved heterogeneity could be rejected, we wanted to decide which of the alternative models that gave the best fit to data. In the testing we employed the full sample and the explanatory variables described in the previous sections.

Table A Model selection

	No unobs. heterogeneity (Model 1)	Normal distribution (Model 2)	Gamma distribution (Model 3)	Discrete distribution (Model 4)	Discrete distribution (Model 5)
Rho		0.60 (0.06)			
Gamma variance			2.07 (0.32)		
Mass Point 1					
Probability				0.70 (0.04)	0.54(0.03)
Value					
Mass Point 2					
Probability				0.30 (0.04)	0.18(0.03)
Value					
Mass Point 3					
Probability					0.28(0.04)
Value					
Information criteria tests:					
AIC	33708,714	33653,772	33646,990	33656,774	33645,726
BIC	33819,910	33768,057	33761,275	33777,237	33769,278
Log likelihood	-16818,357	-16789,886	-16786,495	-16789,387	-16782,863
LR-test statistic		56.94	63.72		

Standard error in parentheses

The likelihood ratio test statistics showed that models which account for unobserved heterogeneity were to be preferred over the model without heterogeneity. The latter model was clearly rejected when tested against the alternatives. This is also confirmed by the significant value of the Rho (Model 2), Gamma variance (Model 3) and the Mass points (Model 4 and 5). For choosing between the models accounting for unobserved heterogeneity, we employed the Akaike's information criterion (AIC) and Bayesian information criterion (BIC) tests. Based on the results presented in Table A, we preferred the model with gamma distribution (Model 3), which had the lowest BIC value and the second lowest AIC value. Hence, it is this model specification that forms the basis for the results presented in Section 4.

References

Bjartveit K, Foss OP, Gjervig T, Lund-Larsen PG. The cardiovascular disease study in Norwegian counties. Background and organization. Acta Medica Scandinavica Suppl 1979; 634; 1-70.

Blaylock, JR, Blisard, WN. Self-evaluated health status and smoking behaviour. Applied Economics 1992; 24; 429-435.

Chaloupka FJ, Warner K. The economics of smoking, in (eds.) AJ. Culyer AJ., Newhouse, JP. Handbook of Health Economics 2000; Amsterdam, Elsevier, 1539–1627.

Clarke A, Etilé F. Do health changes affect smoking? Evidence from British panel data. Journal of Health Economics 2002; 21; 533-562.

Clarke A, Etilé F. Health changes and smoking: an economic analysis. Substance Use & Misuse 2006; 4; 427-451.

DeCicca P, Kenkel D, Mathios A. Putting out the fire: Will higher taxes reduce the onset of youth smoking? Journal of Political Economy 2002;110; 144-169.

DeCicca P, Kenkel D, Mathios A. Cigarette taxes and the transition from youth to adult smoking: Smoking initiation, cessation and participation. Journal of Health Economics 2008; 27; 904-917.

DeCicca P, Mcleod L. Cigarette taxes and older adult smoking: Evidence from recent large tax increases. Journal of Health Economics 2008; 27; 918-929.

Douglas S. The duration of the smoking habit. Economic Inquiry 1998; 36; 49-64.

Farnworth MG. What motivates daily cigarette smokers to quit? Retrospective annual discrete duration analysis. Substance Use & Misuse 2006; 41; 533-554.

Farrelly MC, Bray JW, Pechacek T, Woollery T. Responses by adults to increases in cigarette prices by sociodemographic characteristics. Southern Economic Journal 2001; 68; 156-165.

Forster M, Jones AM. The role of tobacco taxes in starting and quitting smoking: duration analyses of British data. Journal of the Royal Statistical Society 2001; Series A, Part 3, 164; 517-547.

Gallett CA, List JA. Cigarette demand: a meta-analysis of elasticities. Health Economics 2003; 12; 821-835.

Gorin SS, Heck JE. Meta-Analysis of the efficacy of tobacco counselling by health care providers. Cancer Epidemiology, Biomarkers & Prevention 2004; 13;(12); 2012-2022.

Hanagal DD. Frailty regression models in mixture distributions. Journal of Statistical Planning and Inference 2008; 138; 2462-2468.

Hsieh CR Health risk and the decision to quit smoking. Applied economics 1998; 30; 795-804.

Hyland A, Li Q, Bauer, JE, Giovino, GA. Steger, C. Cummings, KM. Predictors of cessation in a cohort of current and former smokers followed over 13 years. Nicotine & Tobacco Research 2004; 6(3); 363-369.

International Agency for Research on Cancer (IARC), Reversal of risk after quitting smoking. 2007; Handbook of Cancer Prevention: Tobacco Control, vol. 11. World Health Organization 2007.

Jenkins S. Easy Estimation Methods for Discrete-Time Duration Models. Oxford Bulletin of Economics and Statistics 1995; 57; 129-138.

Jones AM. Health, addiction, social interaction and the decision to quit smoking. Journal of Health Economics 1994; 13; 93–110.

Jones AM, Rice N, Bago d'Ulva T, Balia S. Applied health economics. Rutledge Advanced Texts in Economics and Finance, London; 2007.

Kalbfleisch JD, Prentice RL. The Statistical Analysis of Failure Time Data. New York, Wiley; 1980.
Kidd MP, Hopkins S. The hazards of starting and quitting smoking: some Australian evidence. The Economic Record 2004; 80; 177-192.
Lancaster T. The Econometric Analysis of Transition Data. Cambridge, Cambridge University Press; 1990.
Levy DT, Chaloupka F, Gitchell J. The effects of tobacco control policies on smoking rates: A tobacco control scorecard. Journal of Public Health Management Practice 2004; 10(4); 338-353.
Madden D. Tobacco taxes and starting and quitting smoking: does the effect differ by education? Applied Economics 2007; 39(5); 613-637.
Nicolás AL. How important are tobacco prices in the propensity to start and quit smoking? An analysis of smoking histories from the Spanish National Health Survey. Health Economics 2002; 11; 521-535.
Ostbye T, Taylor DH. The effect of smoking on "Years of Healthy Life" lost among middle-aged and older Americans. Health Service Research 2004; 39(3); 531-551.
Ostroff JS, Buckshee N, Mancuso CA, Yankelevitz DF, Henschke CI. Smoking cessation following CT screening for early detection of lung cancer. Preventive Medicine 2001; 33; 613-621.
Ross H, Powell LM, Tauras JA, Chaloupka FJ. New evidence on youth smoking behaviour based on experimental price increases. Contemporary Economic Policy 2005; 23; 195-210.
Smith VK, Taylor DH, Sloan FA, Johnson FR Desvousges WH. Do smokers respond to health shocks? Review of Economics and Statistics 2001; 83; 675-687.
Tauras J. The transition to smoking cessation: Evidence from multiple failure duration analysis. NBER working paper, no. 7412; 1999.
Tauras J. Public policy and smoking cessation among young adults in the United States. Health Policy 2004; 68; 321-332.
Tauras J, Chaloupka F. Determinants of smoking cessation: An analysis of young adult men and women, in Economics of Substance Abuse 2001 edited by M. Grossman and C.-R. Hsieh. Edward Elgar Publishing 2001.
Taylor DH, Hasselblad V, Henley, SJ, Thun, MJ, Sloan FA. Benefits of Smoking Cessation for Longevity. American Journal of Public Health 2002; 92(6); 990-996
U.S. Department of Health and Human Services. The health benefits of smoking cessation: A report of the Surgeon General, Atlanta: U.S. Department of Health and Human Services, Public Health Services, Centers for Disease Control, Center for Chronic Disease Prevention and Health Promotion, Office of Smoking and Health, DHHS Publication No. (CDC) 1990; 90-8416.
van der Aalst CM, van Klaveren RJ, van den Bergh KAM, Willemsen MC, de Koning HJ. The impact of a lung cancer CT screening result on smoking abstinence. European Respiratory Journal 2010; published online December 9.
Verbeek M. A Guide to Modern Econometrics, 2dn edition. Chichester, Wiley; 2004.
Warner KE., Mendez D. Tobacco control policy in developed countries: Yesterday, today, and tomorrow. Nicotine & Tobacco Research 2010; 12; 876-887.
WHO 2009 WHO Report on the Global Tobacco Epidemic, 2009: Implementing smoke-free environments. http://www.who.int/tobacco/mpower/2009/en/index.html (accessed 7 Des. 2010).