Working Paper Series
No. 35/2008

Smoking Attributable Health Care Expenditures in Portugal

Anabela Botelho
Elvira M. Lima
Lígia C. Pinto
Paula V. Veiga

October 2008

Núcleo de Investigação em Microeconomia Aplicada
Universidade do Minho

Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA E DA TECNOLOGIA

__ work in progress __

Smoking Attributable Health Care Expenditures in Portugal

by

Anabela Botelho, Elvira M. Lima, Lígia C. Pinto, Paula Veiga

DRAFT #3: October, 2008

NOT FOR CITATION

Authors' Affiliation: NIMA, Escola Economia Gestão, Universidade Minho, 4710-057, Braga, Portugal. Financial support from FCT under POCI/EGE/55490/2004 is greatly appreciated.

1. Introduction

It is well documented, in studies undertaken in other countries, that the economic burden imposed by substance abuse on individuals and society as a whole is substantial. However, in Portugal, the costs of poor health habits have never been evaluated. Here we use data collected by the Portuguese Ministry of Health to assess the health care costs attributable to smoking in Portugal.

While the economic costs of smoking include other dimensions, namely the losses associated with deaths, health care cost is still essential from a public policy perspective. The arguments for any anti-smoking public policies or for any public actions to recover costs are reinforced if the health-care cost attributable to this habit is found to be large. The literature reports two main methods to determine these costs: the "relative-risk method" (RR), and the "microeconometric method" (ME). The RR method uses epidemiological evidence to assess the relative risk to smokers of developing diseases that the medical literature determines to be linked to smoking as compared to non-smokers, figures which, coupled with prevalence rates of the population, are used to ascertain a "smoking-attributable fraction" (SAF) and apply it to estimate the fraction of health care expenditures attributable to the smoking habit. This method, however, has some important limitations: first, it relies on available epidemiological data that generally allows one to control the SAF for sex and age only; second, the relative risks are generally computed based on mortality rather than morbidity data; third, it limits the analyses to diseases that the medical literature establishes as directly caused by smoking.

The recently developed ME method overcomes each of these limitations. Basically, this method follows a different approach to compute the SAF. In short, micro data on health care expenditures and smoking behaviour is used to (i) predict expenditures based on actual behaviour and observed characteristics, (ii) predict expenditures based on the counter-factual scenario where no one has ever smoked; and, the SAF is computed as the relative difference between these predictions. This method, which has been used to measure the health care costs attributable to smoking in the USA, is however extremely demanding in terms of information and it is unusual to find a data set that meets the requirements for its proper application.

The National Health Surveys undertaken by the Portuguese Ministry of Health in several years contain extensive economic, socio-demographic, behavioural, health condition, and health-care cost and use data on the Portuguese population that makes these data sets extremely unique for the full application of the ME method to estimate these costs. In sum, this paper answers the following main questions:

(i) How much are the health-care costs attributable to smoking in Portugal?

(ii) Are these costs statistically and economically significant?

(iii) Do these costs vary in significant ways between the different regions in Portugal?

(iv) Are health care costs attributable to smoking reduced by governmental programs informing people about its health risks, the meaning of the risks, and the consequences of those risks? And if so, are those cost savings statistically and economically significant?

2. The methodology

The microeconometric method used for estimating the SAF in Portugal follows an approach that has been widely used in the literature in recent years (eg. Bartlett *et al*. (1994), Miller *et al*. (1999), Coller *et al.* (2002), Harrison *et al.* (2003)). This approach, pioneered by Duan *et al.* (1983), consists in estimating two-part models to deal with the mixed nature of the distribution of medical expenses. A mixed distribution is a continuous distribution with a positive mass at one or more points (making it neither strictly discrete nor continuous), a common characteristic of medical expenses data where a high percentage of individuals have no medical expenses. Like in our data (Section 3), this corresponds to cumulative distributions showing a "spike" at zero. The two-part account of the mixed nature of these distributions exploits the fact that the likelihood naturally splits the model in two parts: one part deals with the mass at zero, and the other part deals only with those individuals who have positive medical expenses.

Formally, let Y be the medical expense of an individual, Z a vector of covariates (including measures of smoking behaviour), and γ a vector of parameters. Given Z=z, let I=1 if Y>0, and I=0 if Y=0. Then,

$$f_Y(y;\gamma|z)=\begin{cases}\Pr(I=0|z) \text{ if } y=0\\ f_Y(y;\gamma|I=1,z)\Pr(I=1|z) \text{ if } y>0\\ 0 \text{ if } y<0\end{cases} \qquad (1)$$

For a total of n observations where n_1 observations correspond to those individuals with $y_i=0$ (and, therefore, I=0), and $(n-n_1)$ observations have $y_i>0$ (I=1), the likelihood for the parameter vector γ is:

$$\prod_{i=1}^{n} f_Y(y_i;\gamma|z)=\prod_{i=1}^{n_1}\Pr(I=0|z)\prod_{i=n_1+1}^{n} f_Y(y_i;\gamma|I=1,z)\Pr(I=1|z)=$$
$$=\left[\Pr(I=0|z)\right]^{n_1}\left[\Pr(I=1|z)\right]^{n-n_1}\times\prod_{i=n_1+1}^{n} f_Y(y_i;\gamma|I=1,z) \qquad (2)$$

The likelihood is, therefore, the product of two likelihoods, $\prod_{i=1}^{n} f_Y(y_i;\gamma\,|\,z)=(\text{likelihood 1})\times(\text{likelihood 2})$, where the first likelihood corresponds to the probability that an individual has a positive expense, and the second likelihood corresponds to the probability model for the distribution of *positive* expenses only.

Specifying the appropriate probability distributions of each part of the model allows us to obtain consistent and efficient estimates of the parameter vectors by separately maximizing each likelihood (McDowell (2003)). The dependent variable in the first part of the model is dichotomous in nature (either an individual has an expense or not), and a logit or probit specification is often used to estimate the parameters of this part. The dependent variable in the second part of the model is strictly positive commonly with non-constant variance, and a semi-log specification is the most prevalently used to estimate the parameters of this part; this amounts to apply least-squares for logged dependent variables.

Estimation of the first part of the model is used to predict the probability that an individual incurs positive medical expenditures given his or her personal characteristics, including actual smoking habits. Estimation of the second part of the model is used to predict the natural logarithm of an individual’s medical expense level given his or her personal characteristics, including actual smoking habits, *conditional* on the individual having some medical expenses; in this case, the exponential of the log-scale prediction is used to generate the predicted level of medical expenditures for the individual. The overall predicted medical expense for an individual given his or her actual characteristics is then obtained by simply multiplying the predictions from the two parts of the model. Given the actual vector of covariates Z, let $\hat{p}_A$ stand for the predicted probability of having an expense, and $\hat{\mu}_A$ stand for the predicted level of expenses, given an expense is incurred. The overall predicted medical expense is given by:

$$\hat{y}_A = \hat{p}_A \times \hat{\mu}_A \tag{3}$$

Estimation of the SAF proceeds by undertaking a counter-factual simulation of the two-part model. This consists in assuming that all current or former smokers are no longer smokers and indeed never smoked, which amounts to re-setting the values of all smoking-related covariates to zero for all individuals in the data base, and predicting expected expenditures with the statistical model. Let Z^0 be the vector of covariates for an individual with all the smoking-related variables set to zero, but holding all other characteristics at their actual values. Applying the estimated coefficients from the first part of the model above to Z^0 yields the predicted probability that the individual incurs positive medical expenses in the counter-factual scenario where the individual had never smoked. Denote this predicted probability by $\hat{p}_C$, noticing that $\hat{p}_C = \hat{p}_A$ for an individual who actually had never smoked. Similarly, applying the estimated coefficients from the second part of the model above to Z^0 yields, after a suitable transformation of the log-scale prediction, the predicted level of medical expenses in the counter-factual scenario where the individual had never smoked. Denote this predicted expense level by $\hat{\mu}_C$, noticing that $\hat{\mu}_C = \hat{\mu}_A$ for an individual who actually had never smoked.

The overall predicted medical expense in the non-smoking counter-factual scenario is then given by:

$$\hat{y}_C = \hat{p}_C \times \hat{\mu}_C \tag{4}$$

The difference $\hat{y}_A - \hat{y}_C$ is a smoking attributable expense, and the ratio of this difference to the predicted expenses with the actual values of all the covariates is the smoking-attributable fraction for the individual:

$$SAF = \frac{\hat{y}_A - \hat{y}_C}{\hat{y}_A} \tag{5}$$

An aggregate SAF may then be generated multiplying each individual's SAF by their actual expenses, summing up these values for all the individuals with expenses, and dividing the result by the predicted total of the individual's expenses. The aggregate SAF may be applied to the annual health care expenditures to obtain the monetary value of these expenditures that are attributable to smoking.

3. The data

The National Health Surveys undertaken by the Portuguese Ministry of Health in several years contain detailed information on health care expenditures, and on several socio-demographic, economic, and health-behavioural variables for a large sample of individuals in Portuguese mainland allowing the full application of the ME method to estimate an aggregate SAF for Portugal. The surveys employ recognized best-practice survey methods yielding probabilistic samples representative of the mainland Portuguese population.

Here we first use the National Health Survey undertaken in 1998/1999 (a more recent Survey was undertaken in 2005/2006 which is to this day unavailable to the authors) to estimate the SAF. Previous National Health Surveys were undertaken in 1987 and in 1995/1996. The data from each of these surveys allow us to estimate comparable SAFs for Portugal.

Table 1 lists the number of observations by five main regions of Portugal for each of the Surveys, the percentage of observations with zero expenditure, the mean and standard deviation of positive expenditures, and the percentage of individuals who are current or former smokers.

Table 1-Descriptive Statistics with complete samples

Year NHS and Region	N	Percentage with 0 exp.	Mean Expense	Standard Deviation	Percentage Smokers
1998	48,606	67.8	€ 46	€ 129	24.7
North	14,832	71.7	41	90	21.9
Center	9,631	66.3	44	129	20.1
Lisbon	12,608	62.4	56	178	28.6
Alentejo	5,853	68.9	39	67	26.9
Algarve	5,682	71.3	45	95	29.5
1995	49,718	74.1	€ 35	€ 102	23.0
North	14,512	81.9	40	107	19.2
Center	10,039	70.4	32	70	19.0
Lisbon	13,587	69.1	36	130	27.0
Alentejo	6,351	71.6	29	86	25.6
Algarve	5,229	75.7	36	64	27.2
1987	41,585	76.9	€ 14	€ 48	20.5
North	16,127	79.4	14	61	17.9
Center	7,035	78.5	11	18	16.6
Lisbon	11,647	74.4	14	33	23.3
Alentejo	4,118	75.9	14	69	24.2
Algarve	2,658	69.6	17	32	27.9

The descriptive statistics presented in Table 1 are computed using the total number of observations in each of the National Health Surveys. Estimation of the needed statistical models to compute the SAF, however, requires valid observations for each of the relevant variables included in the models. This means that any observations with missing values for any of the relevant variables are discarded from the analysis. There are a number of approaches to prevent discarding observations from the analysis which entail setting missing values to some non-missing values using reasonable imputation algorithms. The simplest imputation algorithm consists in setting missing values on a variable to the mean value of the non-missing values on that same variable. More sophisticated approaches involve developing imputation models for variables with missing values in terms of variables without any missing values. Here, observations with missing values are simply discarded from the analyses, and the robustness of the estimated SAF is then assessed by comparison with the resulting estimates when imputation algorithms are used (TO BE DONE WHEN THE 2005 DATA COMES IN).

Table 2-Descriptive Statistics with working samples

Year NHS and Region	N	Percentage with 0 exp.	Mean Expense	Standard Deviation	Percentage Smokers
1998	35,983	65.2	€ 47	€ 131	29.9
North	10,709	69.0	41	82	27.6
Centre	7,076	63.8	44	121	23.8
Lisbon	9,729	60.3	58	190	33.5
Alentejo	4,542	65.6	38	50	31.9
Algarve	3,927	69.5	45	92	35.7
1995	31,267	72.5	€ 39	€ 120	30.5
North	9,434	80.4	44	125	26.9
Centre	5,835	67.8	34	80	26.2
Lisbon	9,166	67.1	40	151	34.6
Alentejo	3,524	71.8	32	97	33.4
Algarve	3,308	73.6	40	72	33.4
1987	36,882	76.1	€ 13	€ 47	20.9
North	14,298	78.6	14	63	18.5
Centre	6,473	77.8	11	18	17.1
Lisbon	10,215	73.6	13	29	23.8
Alentejo	3,582	75.0	12	68	25.0
Algarve	2,314	68.6	17	32	27.7

Table 2 lists the descriptive statistics for the samples with non-missing values on all the relevant variables. These samples are referred to as "working samples". The figures show that discarding "missing-value" observations reduces the 1998 NHS sample size by about 25%. Compared to the complete sample, the percentage of observations with zero expenditure is lower, and the percentage of individuals who are current or former smokers is higher at the national level in the 1998 working sample. Formal statistical tests on the equality of proportions using large-sample statistics reveal that these proportions are indeed statistically different at conventional significance levels between the complete and working 1998 samples. The application of a t-test on the equality of the positive expenditure means, however, reveals that these figures are not statistically different for these samples.

Concerning the 1995 NHS, discarding "missing-value" observations reduces the sample by about 37%. Comparable formal statistical testing procedures on the equality of proportions and equality of mean expenditures reveals that the figures are indeed statistically significant different amongst the complete and working 1995 samples. The 1987 sample is reduced by about 11% when the missing values are discarded, and only the proportion of individuals with zero expenditure is statistically significantly different between the complete and working samples.

4. Estimated SAF using the 1998 NHS

Selection of the independent variables in the two-part model is driven by previous research findings, and the information that is available in the data set. These variables (but not their coefficients) are the same in each part of the model. Table 3 contains the descriptive statistics for the independent variables of the 1998 working sample. Using acronyms for the variables, we have:

EVERSMOKER: binary indicator that the individual is a current or former smoker.

CIGSNOW: typical number of cigarettes the individual smokes per day now (at the time of the survey).

CIGSNOW2: squared number of cigarettes the individual smokes daily now (CIGSNOW2= CIGSNOW×CIGSNOW).

YEARSSMOKING: number of years the individual has been smoking up to the present survey time.

YEARSSMOKING2: squared number of years the individual has been smoking up to the present survey time (YEARSSMOKING2= YEARSSMOKING×YEARSSMOKING).

CIGSPAST: typical number of cigarettes the individual smoked per day in the past (for a former smoker).

CIGSPAST2: squared number of cigarettes the individual smoked per day in the past (CIGSPAST2= CIGSPAST×CIGSPAST).

YEARSSMOKED: number of years the individual had been smoking in the past (for a former smoker).

YEARSSMOKED2: squared number of years the individual had been smoking in the past (YEARSSMOKED2= YEARSSMOKED×YEARSSMOKED).

DRINKER_NEVER: binary indicator that the individual never drank alcohol.

DRINKER_OFTENWEEK: binary indicator that the individual often drinks alcohol during the week.

DRINKER_ONCEWEEK: binary indicator that the individual drinks alcohol once during the week.

DRINKER_ONCEMONTH: binary indicator that the individual drinks alcohol once during the month.

DRINKER_RARELY: binary indicator that the individual rarely drinks alcohol in the year.

INC: household income group of the individual (1=very low income group, up to 10=upper income group).

REG: region where the individual lives (1=North, 2=Center; 3=Lisbon and Vale do Tejo, 4=Alentejo, 5=Algarve).

AGE: age of the individual in years.

MALE: binary indicator that the individual is male.

MARRIED: binary indicator that the individual is married.

JOB: binary indicator that the individual currently works.

NFAMILY: number of people in the individual's household.

EXER: binary indicator that the individual gets regular exercise at least once a week.

EDU: number of years of schooling.

BMI: body mass index of the individual, defined in terms of reported height and weight.

BMI2: squared body mass index (BMI2= BMI×BMI).

DISABILITY: binary indicator that the individual has a long-term disability.

HIBLOOD: binary indicator that the individual has high blood pressure.

DIABETES: binary indicator that the individual has diabetes.

ASTHMA: binary indicator that the individual has asthma.

BRONCHITES: binary indicator that the individual has bronchitis.

ALLERGY: binary indicator that the individual has allergy.

HEART: binary indicator that the individual has a heart disease (as identified in the ICD9 (9^{th} Revision of the International Classification of Diseases) codes.

CANCER: binary indicator that the individual has cancer (as identified in the ICD9 (9^{th} Revision of the International Classification of Diseases) codes.

PREGNANCY: binary indicator that the individual reported a pregnancy in the past 3 months.

Table 3 – Independent Variables: Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
eversmoker	35983	.2986966	.4576929	0	1
cigsnow	35983	3.54787	8.941875	0	100
cigsnow2	35983	92.54228	333.0409	0	10000
yearssmoking	35983	4.130673	10.66258	0	80
yearssmoki~2	35983	130.75	453.2962	0	6400
cigspast	35983	2.785121	9.576988	0	100
cigspast2	35983	99.47306	494.8414	0	10000
yearssmoked	35983	2.736876	9.146272	0	72
yearssmoked2	35983	91.14246	386.7421	0	5184
drinker_ne~r	35983	.4003835	.4899829	0	1
drinker_of~k	35983	.3901565	.487792	0	1
drinker_on~k	35983	.0553873	.2287378	0	1
drinker_on~h	35983	.0191479	.1370467	0	1
drinker_ra~y	35983	.1342023	.3408742	0	1
inc	35983	5.147959	2.668179	1	10
reg	35983	2.552622	1.322788	1	5
age	35983	49.2397	18.73738	18	103
male	35983	.4665814	.4988889	0	1
married	35983	.670122	.4701751	0	1
job	35983	.511158	.4998824	0	1
nfamily	35983	3.064669	1.324702	1	11
exer	35983	.0795376	.2705797	0	1
edu	35983	5.760359	4.460109	0	24
bmi	35983	25.45403	4.138109	12.81558	66.66667
bmi2	35983	665.0313	227.5542	164.2392	4444.444
disability	35983	.0977406	.2969677	0	1
hiblood	35983	.2102382	.4074834	0	1
diabetes	35983	.065392	.2472197	0	1
asthma	35983	.0570269	.2318973	0	1
bronchites	35983	.0300975	.1708581	0	1
allergy	35983	.1231693	.328636	0	1
heart	35983	.0149515	.1213605	0	1
cancer	35983	.0036128	.0599989	0	1
pregnancy	35983	.0012228	.0349477	0	1

The probit results for the first part of the model are reported in Table 4. Clearly, the results show that the health-related variables are the strongest predictors of positive medical expenditures. Personal characteristics such as age, income, years of schooling, marital status and sex also play a statistically significant role in determining whether or not the individuals have medical expenses. The region where the individuals live also impacts on the probability of having a medical expenditure. The results indicate that, ceteris paribus, individuals living in the Center and in the Lisbon and Vale do Tejo regions of Portugal are more likely to have medical expenses than those who live in the North. The reverse result is found for those individuals living in the Alentejo and the Algarve regions.

An interesting observation is that none of the smoking-related variables is individually statistically significant at less than the 0.05 significance level. The result of a likelihood ratio test of the hypothesis that the coefficients of the smoking-related variables are jointly zero reveals, however, that the set of smoking-related variables is jointly statistically significant in determining whether or not the individuals have medical expenses. The test statistic for the likelihood ratio test is defined as $-2(L_R-L_{UR})$, where L_R and L_{UR} are the values of the log-likelihood functions for the restricted and unrestricted models (the restricted model sets all the coefficients of the smoking-related variables to zero). The computed test statistic is $\chi^2_{(9)}=80.2$ and, therefore, the null hypothesis is rejected.

Table 4 – Probit Results for probability of positive expenditure

Probit regression		Number of obs	=	35983
		Wald chi2(45)	=	3732.89
		Prob > chi2	=	0.0000
Log pseudolikelihood = -21179.488		Pseudo R2	=	0.0888

ypositive	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
eversmoker	.0603015	.0478667	1.26	0.208	-.0335154	.1541184
cigsnow	.0004366	.0037726	0.12	0.908	-.0069575	.0078308
cigsnow2	.0000338	.0000686	0.49	0.623	-.0001007	.0001683
yearssmoking	-.0003561	.0037424	-0.10	0.924	-.0076911	.0069788
yearssmoki~2	3.62e-06	.0000635	0.06	0.955	-.0001208	.0001281
cigspast	.0001688	.0035195	0.05	0.962	-.0067292	.0070668
cigspast2	.0000229	.0000485	0.47	0.637	-.0000722	.000118
yearssmoked	.007809	.004456	1.75	0.080	-.0009247	.0165427
yearssmoked2	-.0000878	.0000774	-1.14	0.256	-.0002395	.0000638
drinker_ne~r	.4554236	.3466115	1.31	0.189	-.2239226	1.13477
drinker_of~k	.2519963	.3466531	0.73	0.467	-.4274312	.9314239
drinker_on~k	.4420499	.3476227	1.27	0.204	-.2392781	1.123378
drinker_on~h	.537468	.3499007	1.54	0.125	-.1483248	1.223261
drinker_ra~y	.4699912	.346873	1.35	0.175	-.2098675	1.14985
_Iinc_2	-.0044648	.0322748	-0.14	0.890	-.0677221	.0587926
_Iinc_3	.0942936	.0326271	2.89	0.004	.0303457	.1582416
_Iinc_4	.0666547	.0335818	1.98	0.047	.0008355	.1324739
_Iinc_5	.0680443	.0345131	1.97	0.049	.0003998	.1356888
_Iinc_6	.1078591	.0353374	3.05	0.002	.0385991	.1771191
_Iinc_7	.0938364	.0371356	2.53	0.012	.021052	.1666208
_Iinc_8	.0914544	.0374503	2.44	0.015	.0180532	.1648556
_Iinc_9	.1108664	.0414235	2.68	0.007	.0296779	.1920549
_Iinc_10	.1021877	.0433632	2.36	0.018	.0171973	.1871781
_Ireg_2	.0550704	.0208745	2.64	0.008	.0141571	.0959838
_Ireg_3	.1056781	.0195871	5.40	0.000	.0672882	.144068
_Ireg_4	-.0617326	.0245102	-2.52	0.012	-.1097717	-.0136936
_Ireg_5	-.1101191	.0260171	-4.23	0.000	-.1611117	-.0591265
age	.0094832	.0006213	15.26	0.000	.0082655	.010701
male	-.2451413	.0179245	-13.68	0.000	-.2802728	-.2100098
married	.1135828	.0171577	6.62	0.000	.0799544	.1472112
job	-.1497198	.0168576	-8.88	0.000	-.18276	-.1166795
nfamily	-.0432857	.0066739	-6.49	0.000	-.0563663	-.0302052
exer	.0352614	.0284426	1.24	0.215	-.0204851	.0910079
edu	.016003	.0023492	6.81	0.000	.0113985	.0206074
bmi	-.0156932	.0130221	-1.21	0.228	-.041216	.0098296
bmi2	.0003502	.0002355	1.49	0.137	-.0001114	.0008119
disability	.2058204	.0251554	8.18	0.000	.1565168	.2551241
hiblood	.3517697	.0184582	19.06	0.000	.3155923	.3879471
diabetes	.2174164	.0287914	7.55	0.000	.1609863	.2738465
asthma	.2306119	.0308274	7.48	0.000	.1701913	.2910324
bronchites	.2081459	.0416613	5.00	0.000	.1264913	.2898005
allergy	.2920898	.0215353	13.56	0.000	.2498815	.3342982
heart	.5758767	.0605009	9.52	0.000	.4572972	.6944562
cancer	.3857404	.1144821	3.37	0.001	.1613595	.6101212
pregnancy	.5790264	.1945962	2.98	0.003	.1976248	.9604279
_cons	-1.256291	.3900596	-3.22	0.001	-2.020793	-.4917877

The regression results for the second part of the model are reported in Table 5. Again, the smoking-related variables are not individually statistically significant, but the result of a standard F test on the joint significance of the set of smoking-related variables clearly indicates that they constitute an important influence on the level of expenditures ($F_{(9,\ 12461)}$=3.46). In line with the probit results, the region where the individuals live also impacts the level of expenditures in a statistically significant manner.

Table 5 – Regression Results for level of expenditure

Linear regression

Number of obs = 12507
F(45, 12461) = 12.01
Prob > F = 0.0000
R-squared = 0.0407
Root MSE = 1.105

lnY	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
eversmoker	-.0587055	.0753354	-0.78	0.436	-.2063745	.0889636
cigsnow	.0098678	.0061145	1.61	0.107	-.0021175	.0218531
cigsnow2	-.000019	.0001068	-0.18	0.859	-.0002284	.0001904
yearssmoking	-.0064644	.0060728	-1.06	0.287	-.018368	.0054392
yearssmoki~2	.0000591	.0000991	0.60	0.551	-.0001352	.0002533
cigspast	.0016777	.0044224	0.38	0.704	-.0069909	.0103463
cigspast2	.0000566	.0000534	1.06	0.290	-.0000482	.0001613
yearssmoked	-.0055115	.0062604	-0.88	0.379	-.0177829	.0067598
yearssmoked2	.0000636	.0001042	0.61	0.541	-.0001407	.000268
drinker_ne~r	-1.056339	.8172928	-1.29	0.196	-2.658359	.5456806
drinker_of~k	-1.253686	.8173337	-1.53	0.125	-2.855786	.3484145
drinker_on~k	-1.175543	.8185145	-1.44	0.151	-2.779957	.4288721
drinker_on~h	-.9529191	.8213476	-1.16	0.246	-2.562887	.657049
drinker_ra~y	-1.175232	.8176065	-1.44	0.151	-2.777867	.4274035
_Iinc_2	.0003214	.0393946	0.01	0.993	-.0768982	.077541
_Iinc_3	.1092089	.041029	2.66	0.008	.0287857	.1896321
_Iinc_4	.1288407	.0441512	2.92	0.004	.0422975	.2153839
_Iinc_5	.1318206	.0458583	2.87	0.004	.0419313	.22171
_Iinc_6	.1705546	.0473374	3.60	0.000	.077766	.2633431
_Iinc_7	.1399813	.051038	2.74	0.006	.0399388	.2400237
_Iinc_8	.2435668	.0510698	4.77	0.000	.143462	.3436716
_Iinc_9	.2823666	.0614701	4.59	0.000	.1618756	.4028576
_Iinc_10	.3333961	.0625431	5.33	0.000	.2108019	.4559903
_Ireg_2	-.0459184	.0294455	-1.56	0.119	-.1036361	.0117993
_Ireg_3	.0577095	.0274831	2.10	0.036	.0038385	.1115805
_Ireg_4	-.0583774	.0335884	-1.74	0.082	-.1242158	.0074611
_Ireg_5	-.0757958	.0390614	-1.94	0.052	-.1523621	.0007706
age	.0035965	.0009109	3.95	0.000	.0018111	.0053819
male	-.0094792	.0266628	-0.36	0.722	-.0617423	.042784
married	.0602835	.0238064	2.53	0.011	.0136192	.1069478
job	-.083486	.0249056	-3.35	0.001	-.1323048	-.0346673
nfamily	-.0189615	.0097686	-1.94	0.052	-.0381095	.0001866
exer	-.0440723	.0478762	-0.92	0.357	-.1379169	.0497724
edu	.0229516	.0035799	6.41	0.000	.0159344	.0299688
bmi	.0202816	.015193	1.33	0.182	-.0094991	.0500623
bmi2	-.0003486	.0002641	-1.32	0.187	-.0008662	.000169
disability	.1188628	.0279755	4.25	0.000	.0640265	.1736991
hiblood	.1468921	.0218075	6.74	0.000	.1041461	.1896381
diabetes	.1232723	.0297028	4.15	0.000	.0650502	.1814944
asthma	.1025288	.0336326	3.05	0.002	.0366037	.1684538
bronchites	.0347742	.042765	0.81	0.416	-.0490519	.1186002
allergy	.1435604	.0263968	5.44	0.000	.0918185	.1953023
heart	.0991384	.0537042	1.85	0.065	-.0061301	.204407
cancer	.1774364	.1403862	1.26	0.206	-.0977423	.4526151
pregnancy	.5192946	.2120483	2.45	0.014	.1036471	.934942
_cons	3.498856	.8480542	4.13	0.000	1.836538	5.161173

Having estimated the two-part model, and generated the relevant actual and counter-factual predictions, each individual's SAF is computed using equation (5). Computation of the aggregate SAF is then a relatively straightforward exercise. In this application, the estimated aggregate SAF is 13.77%. Thus, according to this estimate, about 13.77% of the health care expenditures in Portugal are attributable to smoking, a figure that is clearly significant in economic terms.

5. Estimated SAF using the 1995 NHS

The same logic, statistical procedures and the same definition of covariates as set forth above are used to estimate the aggregate SAF using the 1995 data base. Table 6 contains the descriptive statistics for the independent variables of the 1995 working sample.

Table 6 – Independent Variables: Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
eversmoker	31267	.3045703	.4602326	0	1
cigsnow	31267	3.887965	9.3698	0	100
cigsnow2	31267	102.9066	360.4383	0	10000
yearssmoking	31267	4.13551	10.39181	0	78
yearssmoki~2	31267	125.0888	426.7429	0	6084
cigspast	31267	2.542713	9.050757	0	100
cigspast2	31267	88.37896	457.3998	0	10000
yearssmoked	31267	2.369783	8.29195	0	73
yearssmoked2	31267	74.3701	334.5901	0	5329
drinker_ne~r	31267	.3589727	.4797069	0	1
drinker_of~k	31267	.4212428	.4937662	0	1
drinker_on~k	31267	.0631337	.2432071	0	1
drinker_on~h	31267	.0170787	.1295668	0	1
drinker_ra~y	31267	.1395721	.346548	0	1
inc	31267	6.574408	2.399557	1	10
reg	31267	2.534237	1.308638	1	5
age	31267	45.17686	17.36424	18	101
male	31267	.4903253	.4999144	0	1
married	31267	.6877858	.4634041	0	1
job	31267	.5631177	.4960081	0	1
exer	31267	.0902549	.2865512	0	1
edu	31267	6.233313	3.934915	0	24
bmi	31267	25.13194	3.952267	12.90323	61.63709
bmi2	31267	647.2343	213.205	166.4932	3799.131
disability	31267	.0450635	.2074467	0	1
hiblood	31267	.1646784	.370896	0	1
diabetes	31267	.0490613	.2159995	0	1
asthma	31267	.0272172	.1627184	0	1
bronchites	31267	.0503726	.2187161	0	1
allergy	31267	.1193271	.3241782	0	1
heart	31267	.0121534	.1095722	0	1
cancer	31267	.003742	.061058	0	1

The probit results for the first part of the model using the 1995 working sample are reported in Table 7. Again, the results show that the health-related variables are the strongest predictors of positive medical expenditures. The region where the individuals live also impacts on the probability of having a medical expenditure. The results indicate that, ceteris paribus, individuals living in the Center, in Lisbon and Vale do Tejo, in Alentejo and in the Algarve regions are more likely to have medical expenses than those who live in the North.

In line with the results previously found, none of the smoking-related variables is individually statistically significant at less than the 0.05 significance level. The result of a likelihood ratio test of the hypothesis that the coefficients of the smoking-related variables are jointly zero reveals, however, that the set of smoking-related variables is jointly statistically significant in determining whether or not the individuals have medical expenses. The test statistic for the likelihood ratio test is defined as $-2(L_R-L_{UR})$, where L_R and L_{UR} are the values of the log-likelihood functions for the restricted and unrestricted models (the restricted model sets all the coefficients of the smoking-related variables to zero). The computed test statistic is $\chi^2_{(9)}=56.0$ and, therefore, the null hypothesis is rejected.

Table 7 – Probit Results for probability of positive expenditure

```
Probit regression                                 Number of obs   =      31267
                                                  Wald chi2(42)   =    2678.23
                                                  Prob > chi2     =     0.0000
Log pseudolikelihood = -16938.096                 Pseudo R2       =     0.0796
```

yPositive	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
eversmoker	.0204189	.0515298	0.40	0.692	-.0805777	.1214154
cigsnow	.0009938	.003953	0.25	0.802	-.006754	.0087415
cigsnow2	-9.86e-06	.0000683	-0.14	0.885	-.0001437	.000124
yearssmoking	-.0011373	.0041713	-0.27	0.785	-.0093129	.0070382
yearssmoki~2	.0000468	.0000734	0.64	0.524	-.0000971	.0001907
cigspast	.0000273	.003862	0.01	0.994	-.007542	.0075966
cigspast2	.0000584	.0000533	1.10	0.273	-.0000459	.0001628
yearssmoked	.0046583	.0051083	0.91	0.362	-.0053537	.0146703
yearssmoked2	-5.41e-06	.0000923	-0.06	0.953	-.0001862	.0001754
drinker_ne~r	.1422976	.035214	4.04	0.000	.0732794	.2113157
drinker_of~k	-.0702266	.0348053	-2.02	0.044	-.1384438	-.0020095
drinker_on~h	.0700212	.0669914	1.05	0.296	-.0612795	.2013218
drinker_ra~y	.1265667	.0382052	3.31	0.001	.0516858	.2014475
_Iinc_2	-.0137896	.053994	-0.26	0.798	-.1196159	.0920367
_Iinc_3	.0062944	.0513263	0.12	0.902	-.0943034	.1068921
_Iinc_4	.0463178	.0454284	1.02	0.308	-.0427202	.1353558
_Iinc_5	.0812233	.0431109	1.88	0.060	-.0032725	.165719
_Iinc_6	.0891192	.0414662	2.15	0.032	.0078469	.1703915
_Iinc_7	.0686049	.0400562	1.71	0.087	-.0099037	.1471136
_Iinc_8	.1017327	.0395893	2.57	0.010	.0241391	.1793263
_Iinc_9	.0472169	.0432645	1.09	0.275	-.0375801	.1320138
_Iinc_10	.0866651	.0474627	1.83	0.068	-.0063601	.1796904
_Ireg_2	.3108555	.023717	13.11	0.000	.2643711	.3573399
_Ireg_3	.2828147	.0215881	13.10	0.000	.2405028	.3251265
_Ireg_4	.1483837	.0282036	5.26	0.000	.0931057	.2036618
_Ireg_5	.1058499	.0293689	3.60	0.000	.0482879	.1634119
age	.0092601	.000636	14.56	0.000	.0080135	.0105067
male	-.2300114	.0198183	-11.61	0.000	-.2688546	-.1911682
married	.0379806	.0194061	1.96	0.050	-.0000546	.0760158
job	-.0839737	.0179533	-4.68	0.000	-.1191616	-.0487858

exer	.0548095	.0292674	1.87	0.061	-.0025536	.1121726
edu	.0102712	.0026975	3.81	0.000	.0049842	.0155582
bmi	.006232	.0146478	0.43	0.671	-.0224772	.0349411
bmi2	-.0000369	.0002669	-0.14	0.890	-.0005601	.0004863
disability	.2490485	.037593	6.62	0.000	.1753676	.3227294
hiblood	.3337154	.0217759	15.33	0.000	.2910355	.3763953
diabetes	.2216985	.0354063	6.26	0.000	.1523035	.2910936
asthma	.2354605	.0474312	4.96	0.000	.142497	.3284239
bronchites	.1717484	.0358541	4.79	0.000	.1014756	.2420211
allergy	.3256736	.0235679	13.82	0.000	.2794813	.3718658
heart	.3476422	.0694221	5.01	0.000	.2115775	.483707
cancer	.2520344	.1218874	2.07	0.039	.0131396	.4909293
_cons	-1.569579	.2047961	-7.66	0.000	-1.970972	-1.168186

The regression results for the second part of the model using the 1995 working sample are reported in Table 8. Again, many of the smoking-related variables are not individually statistically significant, and the result of a standard F test on the joint significance of the set of smoking-related variables indicates that they do not constitute an important influence on the level of expenditures ($F_{(9, 8611)}$=1.19).

Table 8 – Regression Results for level of expenditure

Linear regression

Number of obs = 8611
F(42, 8568) = 8.40
Prob > F = 0.0000
R-squared = 0.0393
Root MSE = 1.1696

lnY	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
eversmoker	-.0128187	.0897841	-0.14	0.886	-.1888171	.1631797
cigsnow	-.0160162	.0080043	-2.00	0.045	-.0317065	-.0003259
cigsnow2	.0003049	.0001635	1.86	0.062	-.0000156	.0006254
yearssmoking	.0149389	.0069906	2.14	0.033	.0012357	.0286421
yearssmoki~2	-.0002622	.0001177	-2.23	0.026	-.0004929	-.0000314
cigspast	-.0017553	.006144	-0.29	0.775	-.0137991	.0102885
cigspast2	.0000382	.0000848	0.45	0.652	-.000128	.0002044
yearssmoked	.0037297	.0075024	0.50	0.619	-.0109769	.0184363
yearssmoked2	-.0000463	.0001265	-0.37	0.714	-.0002943	.0002017
drinker_ne~r	.072149	.0958732	0.75	0.452	-.1157857	.2600836
drinker_of~k	-.1425068	.0976056	-1.46	0.144	-.3338373	.0488236
drinker_on~k	-.2017265	.1100752	-1.83	0.067	-.4175003	.0140474
drinker_on~h	(dropped)					
drinker_ra~y	-.0468469	.0994177	-0.47	0.638	-.2417296	.1480358
_Iinc_2	.0572032	.0782772	0.73	0.465	-.0962391	.2106455
_Iinc_3	.0810361	.0711089	1.14	0.254	-.0583544	.2204266
_Iinc_4	.0126106	.0618181	0.20	0.838	-.1085679	.133789
_Iinc_5	.0667879	.0617279	1.08	0.279	-.0542136	.1877895
_Iinc_6	.104758	.05911	1.77	0.076	-.0111118	.2206278
_Iinc_7	.0301402	.0578756	0.52	0.603	-.0833098	.1435902
_Iinc_8	.1439183	.0568743	2.53	0.011	.032431	.2554055
_Iinc_9	.1520109	.0663341	2.29	0.022	.0219801	.2820417
_Iinc_10	.3180431	.0740295	4.30	0.000	.1729273	.4631588
_Ireg_2	-.2241532	.0398921	-5.62	0.000	-.3023514	-.1459551
_Ireg_3	-.1722214	.0369347	-4.66	0.000	-.2446224	-.0998205
_Ireg_4	-.2346215	.0448712	-5.23	0.000	-.3225799	-.1466632
_Ireg_5	-.0558863	.0501006	-1.12	0.265	-.1540955	.0423229
age	.0029499	.0010332	2.86	0.004	.0009246	.0049753
male	.0585252	.0334008	1.75	0.080	-.0069483	.1239988
married	.0786711	.0304158	2.59	0.010	.0190488	.1382935
job	-.1096467	.0304369	-3.60	0.000	-.1693104	-.0499829
exer	-.074522	.0512233	-1.45	0.146	-.174932	.025888
edu	.0238031	.004687	5.08	0.000	.0146154	.0329908
bmi	-.0353975	.0197164	-1.80	0.073	-.0740464	.0032514
bmi2	.0005777	.0003472	1.66	0.096	-.0001029	.0012583

disability	.2078804	.0470383	4.42	0.000	.115674	.3000867
hiblood	.1271628	.0295284	4.31	0.000	.0692801	.1850455
diabetes	.1203123	.0470617	2.56	0.011	.02806	.2125646
asthma	.039251	.0565673	0.69	0.488	-.0716345	.1501365
bronchites	.0409079	.0464613	0.88	0.379	-.0501675	.1319833
allergy	.1064733	.033299	3.20	0.001	.0411992	.1717474
heart	.1306278	.0797943	1.64	0.102	-.0257882	.2870438
cancer	.3946165	.1513713	2.61	0.009	.0978922	.6913407
_cons	3.093033	.2969778	10.42	0.000	2.510884	3.675181

Having estimated the two-part model, and generated the relevant actual and counter-factual predictions, each individual's SAF is computed using equation (5). Computation of the aggregate SAF is then a relatively straightforward exercise. In this application, the estimated aggregate SAF is 18.55%. Thus, according to this estimate, about 18.55% of the health care expenditures in Portugal are attributable to smoking, a figure that is again clearly significant in economic terms.

6. Estimated SAF using the 1987 NHS

Due to lack of availability in the data set, the number of control variables used in the estimation of the two-part model is reduced. The actual control variables used with the 1987 NHS have, however, the same meanings (ie, were constructed in the same way) as those used with the 1995 and 1995 surveys. Table 9 contains the descriptive statistics for the independent variables of the 1987 working sample.

Table 9 – Independent Variables: Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
eversmoker	36882	.2090179	.4066128	0	1
smokenow	36882	.1378179	.3447134	0	1
smoked	36882	.0712	.2571622	0	1
cigsnow	36882	2.557914	7.577591	0	80
cigspast	36882	1.603221	7.105727	0	80
yearssmoking	36882	2.711838	9.097747	0	77
yearssmoked	36882	1.521365	7.020002	0	70
drinker	36882	.4494062	.4974404	0	1
age	36882	38.33924	23.19144	0	100
male	36882	.4769535	.4994753	0	1
married	36882	.5154547	.4997679	0	1
edu	36882	5.638062	3.176016	1	15

job	36882	.4435768	.496813	0	1
reg	36882	2.271759	1.242566	1	5
inc	36882	4.631473	2.632065	1	10
disability	36882	.1565804	.3634097	0	1
diabetes	36882	.0464183	.2103921	0	1
hiblood	36882	.1724961	.3778162	0	1
heart	36882	.0153191	.1228204	0	1
cancer	36882	.0018166	.0425835	0	1

The probit results for the first part of the model using the 1987 working sample are reported in Table 10. Again, the results show that the health-related and personal characteristics variables are the strongest predictors of positive medical expenditures. The region where the individuals live also impacts on the probability of having a medical expenditure, with those living in the Center and those living in the Alentejo regions (Lisbon and Algarve) being less (more) likely to have medical expenses than those who live in the North.

In line with the results found for the 1995 and 1998 samples, the vast majority of the smoking-related variables is not individually statistically significant at less than the 0.05 significance level. The result of a likelihood ratio test of the hypothesis that the coefficients of the smoking-related variables are jointly zero reveals, however, that the set of smoking-related variables is jointly statistically significant in determining whether or not the individuals have medical expenses. The test statistic for the likelihood ratio test is defined as $-2(L_R-L_{UR})$, where L_R and L_{UR} are the values of the log-likelihood functions for the restricted and unrestricted models (the restricted model sets all the coefficients of the smoking-related variables to zero). The computed test statistic is $\chi^2_{(9)}=58.3$ and, therefore, the null hypothesis is rejected.

Table 10 – Probit Results for probability of positive expenditure

```
Probit regression                                 Number of obs   =      36882
                                                  Wald chi2(33)   =    2888.12
                                                  Prob > chi2     =     0.0000
Log pseudolikelihood = -18740.555                 Pseudo R2       =     0.0759
```

ypositive	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
eversmoker	.0253418	.0525055	0.48	0.629	-.0775671	.1282507
cigsnow	-.0035509	.0043962	-0.81	0.419	-.0121673	.0050655
cigsnow2	.0000265	.000082	0.32	0.747	-.0001343	.0001873
yearssmoking	.0055838	.0038517	1.45	0.147	-.0019653	.0131329
yearssmoki~2	-.000066	.000069	-0.96	0.339	-.0002012	.0000693
cigspast	.0050801	.0045339	1.12	0.263	-.0038062	.0139665
cigspast2	-.0001105	.0000725	-1.52	0.128	-.0002526	.0000316
yearssmoked	.0137186	.0047981	2.86	0.004	.0043146	.0231226
yearssmoked2	-.0002053	.0000894	-2.29	0.022	-.0003806	-.00003
drinker	-.3239806	.0181643	-17.84	0.000	-.359582	-.2883792
_Iinc_2	-.0130876	.02819	-0.46	0.642	-.0683391	.0421638
_Iinc_3	.0473099	.031322	1.51	0.131	-.0140801	.1086999
_Iinc_4	.0743544	.0326262	2.28	0.023	.0104082	.1383007
_Iinc_5	.0808345	.0349947	2.31	0.021	.0122461	.1494229
_Iinc_6	.0573722	.0306665	1.87	0.061	-.0027331	.1174774
_Iinc_7	.0687989	.0358844	1.92	0.055	-.0015332	.139131
_Iinc_8	.159628	.0370971	4.30	0.000	.0869191	.2323369
_Iinc_9	.1327107	.0450522	2.95	0.003	.0444101	.2210114
_Iinc_10	.1204799	.0422393	2.85	0.004	.0376923	.2032674
_Ireg_2	-.0526084	.0220526	-2.39	0.017	-.0958307	-.0093861
_Ireg_3	.0608478	.0190456	3.19	0.001	.0235192	.0981764
_Ireg_4	-.0538411	.027317	-1.97	0.049	-.1073814	-.0003008
_Ireg_5	.1474515	.0311364	4.74	0.000	.0864253	.2084777
age	.0098098	.0005775	16.99	0.000	.008678	.0109417
male	-.1384677	.0175458	-7.89	0.000	-.1728568	-.1040786
married	.1209302	.0195961	6.17	0.000	.0825226	.1593378
job	-.2128655	.0174916	-12.17	0.000	-.2471485	-.1785826
edu	-.0026011	.0028141	-0.92	0.355	-.0081167	.0029144
disability	-.102669	.0290394	-3.54	0.000	-.1595852	-.0457529
hiblood	.2961569	.0207859	14.25	0.000	.2554172	.3368965
diabetes	.3364883	.033258	10.12	0.000	.2713038	.4016728
heart	.537326	.0560441	9.59	0.000	.4274816	.6471704
cancer	.7340148	.1635207	4.49	0.000	.4135202	1.054509
_cons	-1.02547	.0339197	-30.23	0.000	-1.091951	-.9589884

The regression results for the second part of the model using the 1987 working sample are reported in Table 11. Again, many of the smoking-related variables are not individually statistically significant, and the result of a standard F test on the joint significance of the set of smoking-related variables indicates that they do not constitute an important influence on the level of expenditures ($F_{(9, 8812)}$=1.23).

Table 11 – Regression Results for level of expenditure

Linear regression

Number of obs	=	8812
F(33, 8778)	=	14.03
Prob > F	=	0.0000
R-squared	=	0.0489
Root MSE	=	1.1483

lnY	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
eversmoker	.1124516	.0894632	1.26	0.209	-.0629172	.2878204
cigsnow	-.0109046	.0093379	-1.17	0.243	-.0292091	.0073999
cigsnow2	.0003057	.000182	1.68	0.093	-.000051	.0006624
yearssmoking	.0011102	.0071934	0.15	0.877	-.0129906	.0152109
yearssmoki~2	-.0000349	.000118	-0.30	0.767	-.0002662	.0001964
cigspast	.0007781	.0069294	0.11	0.911	-.0128051	.0143613
cigspast2	-.0000288	.0001116	-0.26	0.796	-.0002475	.0001899
yearssmoked	-.0172226	.0067285	-2.56	0.010	-.030412	-.0040332
yearssmoked2	.0003288	.0001175	2.80	0.005	.0000985	.0005592
drinker	-.3094322	.0296307	-10.44	0.000	-.3675154	-.2513491
_Iinc_2	.1065288	.0414934	2.57	0.010	.025192	.1878655
_Iinc_3	.0961993	.0483094	1.99	0.046	.0015015	.1908971
_Iinc_4	.1372856	.0505766	2.71	0.007	.0381437	.2364275
_Iinc_5	.1477184	.0585215	2.52	0.012	.0330026	.2624341
_Iinc_6	.10224	.0503399	2.03	0.042	.003562	.200918
_Iinc_7	.2889901	.060449	4.78	0.000	.170496	.4074843
_Iinc_8	.2599645	.0616464	4.22	0.000	.139123	.3808059
_Iinc_9	.4040044	.0751509	5.38	0.000	.2566911	.5513178
_Iinc_10	.3374804	.0757415	4.46	0.000	.1890093	.4859516
_Ireg_2	-.0599012	.0365804	-1.64	0.102	-.1316074	.011805
_Ireg_3	-.0488229	.0319644	-1.53	0.127	-.1114806	.0138347
_Ireg_4	-.2159063	.044783	-4.82	0.000	-.3036915	-.1281211
_Ireg_5	.144326	.0496522	2.91	0.004	.0469961	.2416558
age	.007285	.0008871	8.21	0.000	.0055461	.009024
male	.0338846	.0298516	1.14	0.256	-.0246315	.0924007
married	.0772949	.0300659	2.57	0.010	.0183586	.1362312
job	-.1224593	.0289113	-4.24	0.000	-.1791323	-.0657863
edu	.0204731	.0048791	4.20	0.000	.0109089	.0300373
disability	-.1676624	.0394735	-4.25	0.000	-.2450397	-.0902851
hiblood	.1569875	.0291522	5.39	0.000	.0998423	.2141327
diabetes	.0850472	.0408394	2.08	0.037	.0049925	.1651019
heart	.1552296	.0527999	2.94	0.003	.0517294	.2587298
cancer	.2375741	.1715666	1.38	0.166	-.0987367	.5738848
_cons	1.386749	.0549006	25.26	0.000	1.279131	1.494367

Having estimated the two-part model, and generated the relevant actual and counter-factual predictions, each individual's SAF is computed using equation (5). Computation of the aggregate SAF is then a relatively straightforward exercise. In this application, the estimated aggregate SAF is 19.32%. Thus, according to this estimate, about 19.32% of the health care expenditures in Portugal are attributable to smoking, a figure that is again clearly significant in economic terms.

7. Which SAF?

The previous estimations yield three different figures for the national SAF: 19.32% using the 1987 NHS; 18.55% using the 1995 NHS; 13.77% using the 1998 NHS. A plausible explanation for these different findings is the extent of the information used in estimating the statistical model. The 1998 NHS is the most complete, allowing for a larger number of controls to be used in the estimations. In fact, estimating the two-part model using the 1998 NHS but including only the same controls used in the estimation of the SAF with the 1987 sample yields an aggregate SAF of 18.03%. Similarly, estimating the two-part model using the 1998 NHS but including only the controls used in the estimation of the SAF with the 1995 sample yields an aggregate SAF of 14.48%. It seems, therefore, clear that the estimated national SAF increases as the number of controls are dropped. Given that the use of statistical techniques in computing the SAF is to get at the *pure* effect of smoking on medical expenditures, it seems reasonable to have more confidence on the estimate obtained when more variables were used to control for other factors, apart from smoking, that might be affecting medical expenditures. MORE TO BE ADDED.

8. Adjusting the SAF for Regional variations

Estimation of the two-part model using the NHS surveys revealed that the region where the individuals live are strong predictors both of the probability of having medical expenses and of the level of expenditures conditional on having one. Thus, it is reasonable to adjust the national estimated SAF for these regional variations. To do so, we first determine how the national SAF estimate with the 1998NHS varies with region of residence by stratifying the SAF according to region. Table 12 lists the resulting estimated SAFs by region, showing that the fraction of health expenditures attributable to smoking varies in substantial ways between the different regions in Portugal.

Table 12 – Calculated Smoking Attributable Fractions by Region

1998	13.77%
North	13.62%
Centre	6.51%
Lisbon	15.86%
Alentejo	12.10%
Algarve	24.77%

Secondly, we determine how the aggregate expenditures for the complete 1998 sample are regionally distributed. Table 13 lists these values.

Table 13 – Aggregate expenditures by Region in the 1998 NHS

Region	Total amount	Percentage
Portugal	€ 702198.70	100
North	166692.60	24
Centre	136538.90	19
Lisbon	259048.50	37
Alentejo	70471.15	10
Algarve	69447.50	10

The adjusted national SAF is then computed as:

$(13.62\% \times 24\%)+(6.51\% \times 19\%)+(15.86\% \times 37\%)+(12.10\% \times 10\%)+(24.77\% \times 10\%)=14\%$.

Notice then that the adjusted national SAF reflects the regional composition of the complete 1998 NHS *in terms of expenditures*.

According to the information available at the OECD web site, the total expenditure on health as a percentage of GDP was 10.2% in year 2005, the most recent year for which there is information at the present time. For the same year, the public expenditure on health was 71.8% out of the total expenditure on health. Information available at the same site and for the same year indicates that the Portuguese GDP was 149123.4 million euros at current prices. Applying the estimated adjusted national SAF to these figures (ie, assuming that the 1998 adjusted SAF is valid for year 2005), we have that 2129.482 million euros were national smoking attributable expenses, and that 1528.968 million euros were public smoking attributable expenses.

9. Simulating the impact of information

Although no information exists that would allow us to assess the impact of information on smoking behaviour, it is possible to simulate the cost savings in terms of heath expenditures due to successful informational programs aiming at reducing smoking prevalence or smoking intensity. Based on a small-scale experiment conducted in 1997 in the US using a convenience sample of college students, Botelho (1998) estimated that the provision of information linking smoking to disease (including the full disclosure of information concerning the alleged misconduct by the smoking industry) decreases the probability of smoking by about 50%. Using this value as a benchmark, we may simulate the cost savings due to informational programs in two ways: one way consists in reducing the number of smokers in the sample by 50%; another way consists in reducing the intensity of smoking by the same 50%.

Table 14 reports the estimated SAFs under these two scenarios. As can be gleaned from the Table, only the simulation reducing the number of ever smokers in the 1998 sample produces an impact in the estimated SAFs. The results appear to indicate that informational programs aiming at just reducing smoking intensity do not produce any results in terms of health care expenditures. Although by a small percentage amount, any informational programs aiming at completely reducing the number of smokers *does have* an impact.

Using the same monetary values for health care expenditures as in Section 8, we obtain as a *conservative* figure the value of 40.408 million euros as the cost savings (in terms of public expenditures only) produced by informational programs that reduce by 50% the number of people in the country who smoke.

Table 14 – Estimated counter-factual SAFs in the 1998 NHS

Region	Reducing Number	Reducing Intensity
National SAF	13.39	13.77
North	13.59	13.62
Centre	6.36	6.51
Lisbon	15.34	15.86
Alentejo	11.57	12.10
Algarve	23.75	24.77
Adjusted National SAF	13.63%	14%

10. Conclusion

The present study is a first attempt to estimate the fraction of health care expenditures attributable to smoking in Portugal using an econometric methodology that is now an accepted standard in the health economics literature. In line with the results in studies undertaken in other countries, our findings suggest that a substantial fraction of the national health care expenditures can be attributed to smoking. Informational campaigns and other public initiatives aiming at preventing people from starting to smoke and helping those who already smoke to quit this highly addictive habit are, therefore, justified in economic terms. Our simulations suggest that substantial cost savings can be achieved even if such schemes are not totally successful.

References

Bartlett, J.; Miller, L.; Rice, D., and Max, W. (1994), "Medical Care Expenditures Attributable to Cigarette Smoking: United States, 1993", *Morbidity and Mortality Weekly Report*, 43, 469-472.

Botelho, A. (1998), "The Quantification of Economic Damages Attributable to Cigarette Smoking: Some Methodological Suggestions", *Proceedings of the 3rd Conference of the Sociedade Portuguesa de Investigação em Economia (SPiE)*, 1998.

Coller, M.; Harrison, G. W. and McInnes, M. (2002), "Evaluating the Tobacco Settlement: Are the Damages Awards Too Much or Not Enough?, *American Journal of Public Health*, 92(6), 984-89.

Duan, N.; Manning, W.; Morris, C., and Newhouse, J. (1983), "A Comparison of Alternative Models for the Demand for Medical Care", *Journal of Business and Economic Statistics*, 1(2), 115-126.

Harrison, G.W.; Feehan, J.P.; Edwards, A.C., and Segovia, J.(2003), "Cigarette Smoking and the Cost of Hospital and Physician Care", *Canadian Public Policy-Analyse de Politiques*, XXIX(1), 1-20.

McDowell, A. (2003), "From the help desk: hurdle models", *Stata Journal*, 3(2), 178-184.

Miller, V.; Ernst C., and Collin, F. (1999). "Smoking-Attributable Medical Care Costs in the USA", *Social Science and Medicine,* 48(3), 375-391.