HEDG
HEALTH, ECONOMETRICS AND DATA GROUP
THE UNIVERSITY of York

WP 12/06

Time and chance happen to them all? Duration modelling versus lifetime incidence of cancer

Daniel Howdon

April 2012

york.ac.uk/res/herc/hedgwp

Time and chance happen to them all?*Duration modelling versus lifetime incidence of cancer.

Daniel Howdon
University of York

March 14, 2012

Abstract

While current work on socioeconomic inequality in cancer looks at lifetime incidence of cancer, it is more informative to consider survival times: healthy time lived without cancer. This paper uses the first wave of, and latest longitudinal follow-up to, the Health and Lifestyle Survey (HALS) to investigate the social gradient in cancer, considering both lifetime incidence and duration models of time-to-cancer. Contrary to previous work on the relationship between circumstances and the development of cancer, notably Deaton (2002), a social gradient in time to cancer is observed, with those in the lowest two social classes developing cancer significantly (at the 5% level of significance) sooner than individuals in the highest social class. This relationship holds after excluding smokers from the sample. No gradient is observed when only lifetime incidence of cancer is considered.

*Ecclesiastes 9:11. "I have seen something else under the sun: The race is not to the swift or the battle to the strong, nor does food come to the wise or wealth to the brilliant or favour to the learned; but time and chance happen to them all."

1 Introduction

While strong evidence exists regarding a social gradient in lifespan overall and illnesses such as cardiovascular disease, the existence of a social gradient in cancer is more controversial. Deaton (2002) argues that the Whitehall Studies (Marmot et al., 1978, 1991) show no social gradient in any cancer apart from lung cancer, the gradient in which is entirely explained by differences in smoking between the occupational grades. Despite finding social gradients in health overall and in many diseases, Wilkinson and Pickett (2010) find no social gradient in breast cancer, and 'only small class differences' in prostate cancer. Much of the attention in this area has focused on incidence of cancer rather than survival time to cancer (for instance, Singh et al. (2003); Banks et al. (2006); Dalstra et al. (2005)): the differences in results obtained by the use of such approaches are highlighted in this paper.

No study has, as yet, exploited the Health and Lifestyle Survey (HALS) dataset and the subsequent cancer follow-ups for the purposes of carrying out such an investigation. Hiatt and Breen (2008) identify the question of why social determinants are correlated with the development of cancer – because they are correlated with already-understood risk factors for cancer or biological factors, or because they are inherently causative of cancer – as 'a key question in cancer research'. Further, the authors call for the investigation of such factors to encourage 'a more complete understanding of the causes of cancer'. While the possibility of causal analysis is limited by the available dataset, this paper provides evidence of a link between social class and cancer, independent of differences in smoking behaviours, and goes some way towards addressing these concerns.

The importance of smoking in this analysis is clear: a link between smoking and ill-health in general, and cancer specifically, is uncontroversial. The risks of smoking have been well-explored since the link between smoking and lung cancer was made by Doll and Hill (1954). Smoking has been associated with a greater propensity to develop various cancers and other diseases (for instance, deaths from lung cancer are estimated to occur with between 10.8 and 24.9 times the frequency in smokers as in non-smokers (Doll, 1998)) and is estimated to be responsible for approximately 30% of all cancer deaths in developed countries, as well as causing deaths from respiratory, circulatory and other problems (Department of Health and Human Services, 1989; Jones et al., 2007; Peto et al., 2006; Vineis et al., 2004). Using Health Survey for England data

from 1998 to 2006, Vallejo-Torres and Morris (2010) estimate that 2.3% of all socioeconomic (income-related) inequality in health observed was due to smoking. Successive reports by the US Surgeon General (Department of Health and Human Services, 1989, 2004, 2010) have examined the evidence linking smoking with mortality and diseases including cancer, increasingly making stronger links over time, with 30 diseases listed in the 2004 report for which evidence was 'sufficient to infer a causal relationship'. Doll (1998) provides a useful summary of the history of mounting evidence regarding the links between smoking and ill-health. Given this, along with the disproportionate levels of smoking among those in lower social classes and the possibility of this acting as a confounding factor when estimating relationships between social class and ill-health, three different models are estimated, each treating smoking variables differently. A first model includes smokers in the sample, but excludes variables for smoking behaviours in the regression; a second model includes smokers in the sample and includes variables for smoking behaviours in the regression; a third and final model excludes smokers completely. Further, all individuals aged under 45 are excluded.

In order to make any analysis of equality of opportunity in health, it must be first established how far each separate factor affects an individual's health – as discussed above, an individual's time alive without being registered for cancer is likely to be a function of both circumstances and choices. Such (non-duration) analysis is conducted by Balia and Jones (2008), using HALS1 and the survey's mortality follow-up of May 2003, finding that lifestyles make an important contribution to observed socio-economic gradients in health, with smoking making a particular strong contribution. This analysis was extended to socioeconomic gradients in smoking prevalence and duration by Balia and Jones (2011). In this paper, we use the most recent HALS follow-up on mortality and cancer registration: that of July 2009. A correlation, significant at the 5% level, between social class and accelerated time to cancer is observed.

2 Data

This paper uses the Health and Lifestyle Survey (HALS), conducted between 1984 and 1985, together with the most recent longitudinal follow-up (that of July 2009). HALS contains data on lifestyle, behaviours (such as smoking and alcohol consumption) and circumstances of a large

cross-section of a representative sample of individuals in Great Britain[1] (Cox et al., 1993). Data collection consisted of a one-hour face-to-face interview to collect information on individuals' health, lifestyles, and socio-demographic data. It involved a visit from a nurse to collect information on physiological and cognitive function, and a self-completed questionnaire to gather information regarding psychiatric health and personality (Cox et al., 1993; Jones et al., 2007). Details of individuals' diagnoses with cancer and information relating to individuals' deaths (such as date and cause of death) were subsequently provided to the HALS team. Such data, including details from death certificates and cancer diagnoses, are used, correct to the beginning of July 2009 – the seventh deaths revision and fourth cancer revision (University of Cambridge Clinical School, 2009). 9,003 individuals were initially entered into the study of whom, as of this revision, the status of 97.8% had been flagged on the NHS's Central Register at the Office for National Statistics: 2,883 individuals have been flagged as dead and 1,468 coded for cancer.

The original HALS dataset contains 9,003 observations: after cleaning, this is reduced to 8,213. This cleaning procedure was particularly assiduous for smoking-related variables, and required the removing of individuals who did not report their smoking status, individuals who were reported to be current smokers but did not report age at starting, individuals who were reported to have started smoking at a date later than their date of interview, ex-smokers who did not declare when they quit or the age at which they began to smoke, current and ex-smokers who did not provide an average number of cigarettes smoked. Data was also cleaned, and records for individuals deleted, where they were reported to have died on or before their date of interview, registered for cancer on or before their date of interview, or where any other variable included in the regression was missing for the individual. The exclusion of individuals aged under 45 at the time of HALS further reduces the sample size to 3,800, and the exclusion of smokers (in the final model) reduces the sample to 1,397 individuals.

Descriptive statistics for the full survey, and the full survey after cleaning, are presented in Table 1[2].

Given that this survey took a cross-section of individuals living in Great Britain at one point in time, there is the possibility for, and such cases exist where, individuals had been diagnosed with cancer prior to their being interviewed for HALS. While the exclusion of those living with

[1] That is the United Kingdom, not including Northern Ireland.

[2] It should be noted that some of these descriptive statistics – particularly those smoking-related – are potentially unreliable pre-cleaning.

		Mean			
Description	Label	n=9003	n=8213	n=3800	n=1397
HALS smoker	start	0.346	0.340	0.304	-
Pack-years (HALS1 smoker)	packyrss	7.082	7.414	9.845	-
Pack-years squared (HALS1 smoker)	packyrs2s	0.026	0.027	0.043	-
HALS ex-smoker	quit	0.248	0.254	0.329	-
Pack-years (HALS1 ex-smoker)	packyrsq	4.881	5.098	8.861	-
Pack-years squared (HALS1 ex-smoker)	packyrs2q	0.024	0.026	0.051	-
Non-prudent drinker of alcohol	NPAD	0.121	0.124	0.087	0.039
Eats meat 3+ times per week	redmeat3	0.452	0.451	0.527	0.510
Highest educational qualification: degree	lhqdeg	0.046	0.047	0.027	0.037
Highest educational qualification: other	lhqoth	0.006	0.006	0.006	0.006
Highest educational qualification: A-level or equivalent	lhqA	0.044	0.041	0.031	0.029
Highest educational qualification: O-level or equivalent	lhqO	0.143	0.148	0.072	0.080
Long-term unemployed (one year or more)	ltunemp	0.03	0.029	0.020	0.007
Not working due to sickness	sick	0.021	0.020	0.038	0.020
Retired	retd	0	0.209	0.451	0.481
Single	single	0.17	0.166	0.066	0.077
Separated or divorced	sepdiv	0.06	0.061	0.050	0.048
Widowed	widowed	0.087	0.080	0.166	0.204
Male	male	0.434	0.443	0.451	0.282
Social class 2 or 3	sc23	0.655	0.681	0.652	0.674
Social class 4 or 5	sc45	0.277	0.289	0.319	0.293
1920s birth	bc20	0.156	0.155	0.335	0.268
1930s birth	bc30	0.153	0.157	0.307	0.324
1940s birth	bc40	0.193	0.201	0.000	0.001
1930s or 1940s birth	bc3040	0.347	0.358	0.307	0.325
1950s birth	bc50	0.189	0.195	-	-
1960s birth	bc60	0.13	0.126	-	-

Table 1: Descriptive statistics, full survey and cleaned full survey

cancer in 1985/6 does mean that the sample is necessarily less representative of the population, this avoids the problem of the inclusion of such individuals with apparently negative survival times, which cannot be modelled using the distributions employed here.

2.1 Clarifications

It must be borne in mind there are delays involved in the registration of deaths and developing cancer, and that delays are not uniform in all cases. The latest HALS follow-up manual (University of Cambridge Clinical School, 2009) suggests that cancer registrations tend to be slower to reach the NHS Central Register than death notifications (although such registrations are 'probably' complete up to the end of 2007), and that missing cases will exist due to 'patchy' returns from regional registries. Spikes are recorded in 2008 and 2009 for individuals who died with cancer present without ever being registered as developing such a disease (Table 2), suggesting that some late returns may exist for this revision[3]. Comparison of the previous HALS follow-up (to April 2005) with data held in this latest follow-up shows, however, that no cancer registrations were late – i.e. were included in the July 2009 follow-up with a date of April 2005 or earlier[4]. Furthermore, the age at the time of an individual's first cancer registration is not the same as the age of the individual first developing cancer. Diagnosis of cancer does not immediately take place upon the individual developing the disease, nor does it occur at the same stage of development of the cancer across individuals, or over time. In particular, the stage at diagnosis has varied over time, with US National Cancer Institute (2006) showing declines in the rates of late-stage diagnoses of cases of cancers of the cervix, colon, prostate and rectum between 1980 and 2006.

If a social gradient exists in smoking behaviours, it may well be the case that excluding those who claimed to have ever smoked in 1985/6 does not fully remove the effect of any potential correlation between smoking and cancer: if individuals in lower social classes are more likely to smoke, they are also more likely to take up smoking after HALS was conducted, and consequently some of the apparent correlation with social class may be explained by differential smoking behaviours post-HALS. However, the probability of smoking take-up drops off precipitously after adolescence, and is almost zero by the time an individual reaches the age of 30

[3]This data is obtained using the Stata icd9 command to search for individuals whose death certificate shows any cancer (codes in the range 140 to 239.99).

[4]Seven death registrations were, however, late by this measure.

Year of death	No. of deaths	Percentage
1985	5	3.42
1986	17	11.64
1987	14	9.59
1988	18	12.33
1989	27	18.49
1990	23	15.75
1991	5	3.42
1992	1	0.68
1993	4	2.74
1994	4	2.74
1995	1	0.68
1996	1	0.68
1997	2	1.37
2000	2	1.37
2001	1	0.68
2002	2	1.37
2006	2	1.37
2007	1	0.68
2008	10	6.85
2009	4	2.74
Total	146	

Table 2: Deaths where cancer is listed on an individual's death certificate, with the individual never registered as developing cancer (full survey)

(Douglas and Hariharan, 1994). Further, dropping individuals who had ever smoked before 1985/6 may cause a masking of the initial selection effect into smoking: if individuals in lower social classes have some prior information that they are likely to die or develop cancer sooner and make decisions regarding smoking behaviours on this basis, some of the apparent effect of social class may be due to selection effects.

Variables in the model measured at HALS are effectively assumed to be time-invariant: there is no way to establish how these variables subsequently (and, in most cases, previously) changed. In the case of social class, individuals' experience of being in a particular class, and the effect on their health of being in that class, is not constant over time, even if they remain in that same class, due to changes in, for instance, relative and absolute incomes over time. Being in social class V in 1985, for instance, does not necessarily have the same impact on health as being in social class V in 2010.

A variable for long-term unemployment (those unemployed for a period of one year or more) is included in the model to control for individuals who may have been suffering from only a short spell of worklessness. While correlation between long-term unemployment and ill-health is well-established, evidence differs regarding the direction of causality. Gordo (2006) claims, accounting for endogeneity, that long-term unemployment has a significant and negative effect on the health of individuals (using German data), while Böckerman and Ilmakunnas (2009) (using Finnish data) conversely suggest that individuals with poor health prospects are sorted into unemployment[5].

3 Methods

3.1 Statistical models

This paper assesses the correlation between social class, and other lifestyle and socioeconomic variables, and life years from birth without cancer[6].

Figure 1 illustrates the basic possibilities for different types of individuals in the data. The horizontal axis represents time, with events to the left occurring before events to the right, and examples of subject types appearing on the vertical axis. Date of birth and dates of starting

[5]See Mathers and Schofield (1998) and Böckerman and Ilmakunnas (2009) for a review of the evidence on the relationship and possible direction of causation between unemployment and health.

[6]Analysis is carried out using Stata version 11.0.

and quitting smoking were collected in the initial HALS survey, and date of death in subsequent follow-ups. Using this information, a solid line denotes years alive (survival time in the lifespan models), with a black circle denoting birth and a black square denoting death (failure in the lifespan models). The dashed line beyond July 2009 represents the fact that these observations are right-censored at this point as such individuals' status as alive or dead (or registered cancer sufferers or not) is not known beyond this. Further discussion of this diagram follows, with a key provided within Figure 1.

3.1.1 Cancer registration

Non-smokers only are included in this model (a, b, g, h in the above diagram), and are entered into the model conditional on survival at the time of HALS. While individuals can be, and indeed are, observed to have developed cancer before the survey began, individuals who had developed cancer before HALS are much more likely to have died before the survey took place. Those individuals with pre-existing cancer registrations are dropped from the sample[7]. Individuals who are registered as dead at the time of the most recent follow-up are checked for any appearance of a cancer on their death certificate. Such individuals are treated as failures in this model, with a failure time of their age at death. The dependent variable here is healthy time observed (*cancerage*): i.e. time before an individual is observed to have developed cancer. Individuals who have been registered as developing cancer at the time of the July 2009 HALS follow-up (g and h, of those here included) have a complete spell observed for this model while individuals who have never been registered as developing cancer at this time (a and b) are censored. Maximum likelihood estimation is used to estimate the link between various factors and the associated acceleration of time to failure.

This cancer registration model is clearly problematic in terms of interpretation. While cancer registration, if it occurs, must clearly precede death, death cannot precede cancer registration[8]. Consequently, individuals can be censored in this model for two reasons: that they are not registered as having developed cancer at the time of the follow-up (a), or that they have died without being registered as cancer or having cancer on their death certificate(b). These two

[7]Further, the inclusion of such individuals would lead to some individuals effectively having negative survival times in the left-truncated survival model used here.

[8]However, individuals can have a cancer registration age equal to their age at death, where cancer appears on the death certificate without the disease ever being previously diagnosed.

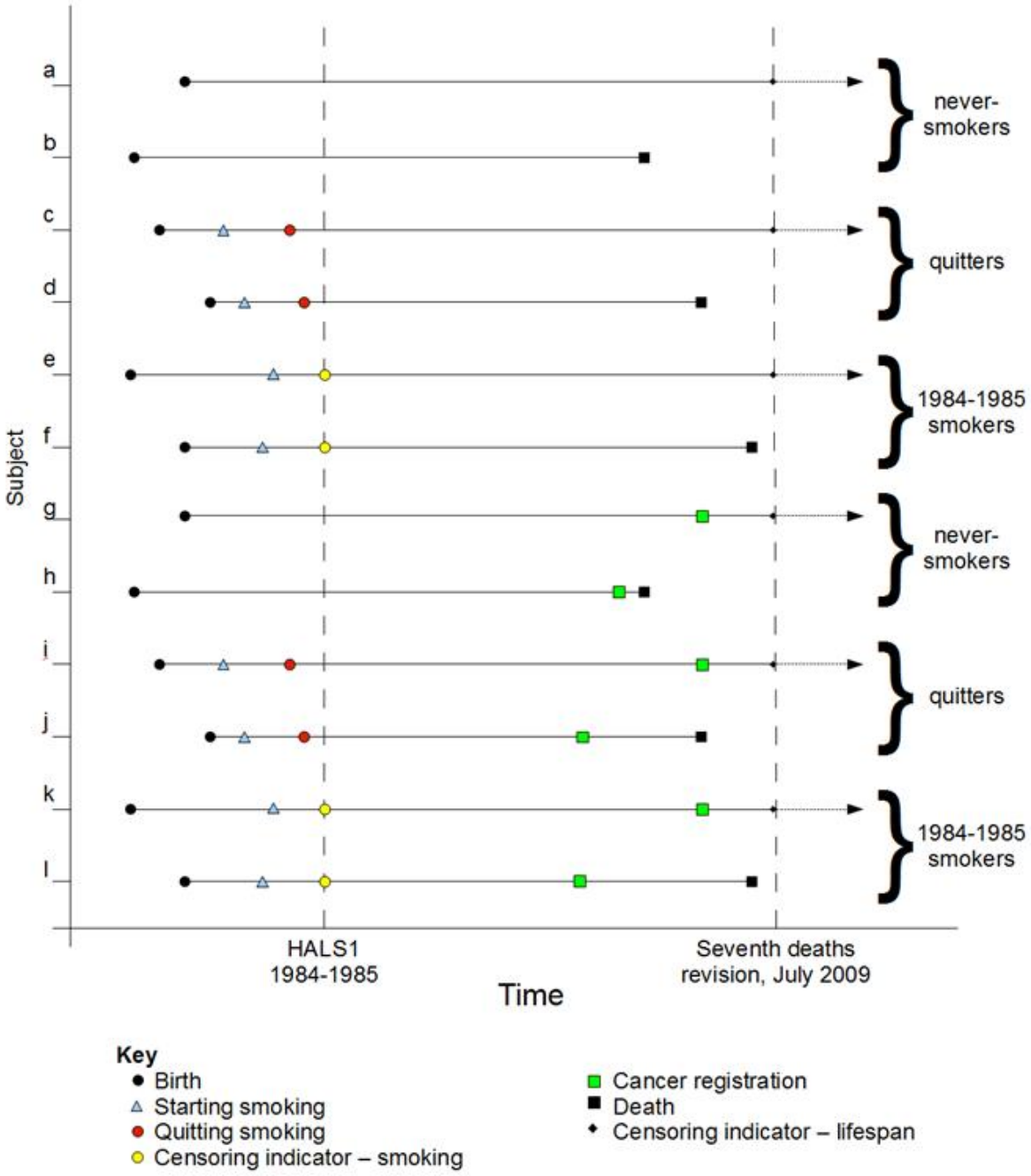

Figure 1: Types of individual

types of censorings clearly differ. While survival (i.e., being alive and not registered as a cancer sufferer) at HALS is plausibly non-informative, death (particularly from certain causes) is potentially informative: for instance, cardiovascular disease and some cancers (such as lung cancer) share risk factors. Death from such diseases is therefore likely to be correlated with cancer registration: those dying from, for instance, CVD are more likely to, absent such a death, have developed cancer. The example of CVD is particularly pertinent given that smoking causes CVD with a relatively short lag and lung cancer with a much longer lag (Cutler et al., 2006). As such, deaths are not accurately characterised as non-informative censorings but, where the cause of death is etiologically similar to cancers or the individual has innate susceptibilities to both the cause of death and cancers (Estève et al., 1994), death is likely to be correlated with the potential for cancer registration absent death. While a formal specification of the joint distributions of survival times for cancers and deaths is required to entirely eliminate any biases, such information is inherently unavailable (Estève et al., 1994; Honoré and Lleras-Muney, 2006). However, if individuals who die sooner are also disproportionately likely to develop cancer absent their death, and if being in a low socioeconomic group is correlated with both accelerated time to cancer and accelerated time to death, the coefficients for social class in this model provide (in absolute terms) lower limits for the true coefficients.

4 Results

Distributions for the hazard of cancer registration (with all smokers excluded) are compared on Akaike Information Criteria and Bayesian Information Criteria scores, which take the log-likelihood generated through maximum likelihood estimation and impose a penalty for the introduction of additional parameters which more flexible distributions include[9]. Details for the final model are given in Table 3. A lognormal distribution is selected for this model – on both BIC and AIC score, it outperforms all other competing distributions [10].

Two preliminary models include smokers in the sample, for which results are shown in Tables 4 and 5. A social gradient – specifically, a significant difference in time to failure between class 1 and all other classes – in observed in cancer registrations when no account is taken of smoking

[9]For instance, the most flexible distribution compared here, the generalised gamma, includes three extra parameters than the least flexible, the exponential distribution. See, for instance, Cox et al. (2007) for more details on these nested distributions.

[10]Results are robust to the use of the more flexible generalised gamma distribution.

Model	Loglikelihood	degrees of freedom	AIC	BIC
Generalised gamma	-273.931	20	587.8621	692.7037
Lognormal	-274.4143	19	586.8285	686.4281
Loglogistic	-275.1758	19	588.3515	687.9511
Gompertz	-279.2357	19	596.4715	696.0711
Weibull	-276.6025	19	591.205	690.8046
Exponential	-299.1227	18	634.2455	728.603

Table 3: Comparison of AIC and BIC scores

behaviours (Table 4), but no such significant gradient is observed when these variables (dummy variables for smoking status, and variables for pack-years and pack-years squared) are included in the estimation, reflecting the disproportionate smoking exhibited among those in lower social classes[11]. This would seem to support the hypothesis that the social gradient in cancer is explained by differences in smoking behaviours. However, to exclude the possibility (subject to the earlier clarification and assumption regarding smoking behaviours post-HALS) of selection into smoking based on an individual's own private information regarding his health expectations (as in, for instance, Balia and Jones (2011)), and to exclude the extent to which the hypothesis of Link and Phelan (1995) is correct, a final model – dropping all smokers – is estimated, and results presented in Table 6.

All smokers are dropped from the sample in this final model and, in such a way, ignoring the effect of smoking altogether, a persistent gradient in social class is observed, as shown in Table 6, below.[12][13]

Socioeconomic inequalities appear to exist in time to cancer registration, even after the exclusion of all smokers from the sample. This gradient appears to be less obvious when the model is estimated on the basis of lifetime incidence only, appearing only significant at the 10% level when estimated in probit form. Being in social class 2, 3, 4 or 5 is significant at the 5% level in reducing time to cancer by approximately 15%, with a slightly greater effect for social classes 4 and 5 than 2 and 3.

[11]For instance, 48% of those in the original HALS sample in unskilled occupations were regular smokers, compared to 22% in the highest social class.

[12]Note that, while the sign attached to many variables here is reversed between probit and lognormal distributions, this is to be expected. A positive coefficient in the probit model indicates an increased probability of the individual being registered as a cancer suffered over his or her lifespan, while a positive coefficient in the lognormal model indicates an *increase* in the expected time before the individual develops cancer.

[13]Only one individual was born in the 1940s in this model, and consequently bc40 is excluded from the probit model.

	Lognormal		Probit	
	Coef.	Std. err.	Coef.	Std. err.
NPAD	-0.071***	0.027	0.182**	0.079
redmeat3	0.01	0.015	-0.034	0.044
lhqdeg	0.097**	0.048	-0.270*	0.15
lhqoth	0.06	0.099	-0.271	0.288
lhqA	0.06	0.044	-0.11	0.131
lhqO	0.016	0.028	0.034	0.085
ltunemp	-0.134**	0.052	0.300**	0.153
sick	-0.047	0.041	-0.07	0.118
retd	0.028	0.024	-0.057	0.07
single	0.004	0.032	-0.085	0.09
sepdiv	0.003	0.034	-0.039	0.104
widowed	0.03	0.025	-0.140**	0.066
male	-0.052***	0.017	0.081*	0.048
sc23	-0.085*	0.047	0.162	0.141
sc45	-0.115**	0.049	0.182	0.145
bc20	0.034	0.026	0.006	0.067
bc3040	0.052	0.035	-0.282***	0.086
Constant	4.555***	0.055	-0.659***	0.164
ln(sigma)	-1.267***	0.053		
N. of cases	3800		3800	

* $p < 0.10, * * p < 0.05, * * *p < 0.01$

Table 4: Lognormal regression, smoking variables excluded

	Lognormal		Probit	
	Coef.	Std. err.	Coef.	Std. err.
start	0.019	0.038	-0.025	0.120
packyrss	-0.006***	0.002	0.011**	0.005
packyrs2s	0.412**	0.206	-0.867	0.586
quit	-0.025	0.025	0.056	0.075
packyrsq	-0.001	0.001	0.000	0.003
packyrs2q	-0.023	0.083	0.134	0.215
NPAD	-0.046*	0.026	0.144*	0.080
redmeat3	0.015	0.015	-0.043	0.044
lhqdeg	0.076*	0.046	-0.235	0.15
lhqoth	0.057	0.095	-0.280	0.290
lhqA	0.057	0.042	-0.108	0.132
lhqO	0.010	0.027	0.045	0.086
ltunemp	-0.107**	0.050	0.251	0.154
sick	-0.028	0.040	-0.104	0.119
retd	0.025	0.023	-0.056	0.07
single	-0.004	0.030	-0.072	0.09
sepdiv	0.011	0.032	-0.056	0.104
widowed	0.029	0.024	-0.141**	0.066
male	-0.033**	0.017	0.049	0.051
sc23	-0.064	0.045	0.13	0.142
sc45	-0.083*	0.047	0.133	0.146
bc20	0.036	0.025	-0.016	0.068
bc3040	0.040	0.033	-0.293***	0.087
Constant	4.570***	0.053	-0.682***	0.165
ln(sigma)	-1.303***	0.052		
N. of cases	3800		3800	

* $p < 0.10, * * p < 0.05, * * * p < 0.01$

Table 5: Loglogistic regression, smoking variables included

	Lognormal		Probit	
	Coef.	Std. err.	Coef.	Std. err.
NPAD	-0.078	0.060	0.328*	0.191
redmeat3	0.039	0.024	-0.113	0.076
lhqdeg	0.010	0.064	-0.091	0.213
lhqoth	0.011	0.148	-0.096	0.468
lhqA	0.009	0.069	0.024	0.221
lhqO	-0.003	0.043	-0.001	0.141
ltunemp	-0.035	0.133	-0.151	0.468
sick	-0.012	0.089	-0.137	0.28
retd	-0.026	0.04	0.081	0.125
single	-0.043	0.047	0.090	0.139
sepdiv	0.009	0.055	-0.092	0.181
widowed	0.030	0.037	-0.137	0.105
male	-0.044	0.028	0.113	0.088
sc23	-0.155**	0.074	0.428*	0.24
sc45	-0.157**	0.077	0.358	0.25
bc20	0.011	0.040	0.000	0.116
bc3040	-0.029	0.051	-0.170	0.148
Constant	4.709***	0.086	-1.066***	0.277
ln(sigma)	-1.301***	0.086		
N. of cases	1397		1397	

* $p < 0.10, * * p < 0.05, * * *p < 0.01$

Table 6: Lifetime incidence (lognormal) and healthy time before cancer (probit) models

5 Conclusions

Most existing studies of cancer incidence concentrate solely on lifetime incidence of types of the disease. Here, such incidence is compared with time to cancer, an innovation made possible by the format of HALS, incorporating a single cross-section in 1985 followed by future waves to provide data on cancer registrations. The derived correlation should be seen as a lower bound on the actual relationship between social class on cancer outcomes. Interpretation of coefficients in the cancer registration model is complicated by the way in which those who do not develop cancer are censored: (at least some) deaths are not non-informative censorings, but are symptomatic of the tendency of the individual to develop cancer, in the absence of death. Under the assumptions discussed earlier, however, these coefficients can be treated as (in absolute terms) lower bounds on their true values.

Further, if the analysis of Link and Phelan (1995) is correct, the total effect of circumstantial factors such as social class, where correlated with particular lifestyles, is likely to be underestimated. Were people in lower social classes not to disproportionately adopt certain unhealthy lifestyles, for instance smoking, they would adopt others. Such dynamic effects are inevitably excluded from this model, and could only be identified using a dynamic analysis which looked at changing behaviours over time, and potential substitutions out of, for instance, smoking into other risky behaviours.

References

Balia, S. and Jones, A. M. (2008), 'Mortality, lifestyle and socio-economic status', *Journal of Health Economics* **27**(1), 1–26.

Balia, S. and Jones, A. M. (2011), 'Catching the habit: a study of inequality of opportunity in smoking-related mortality', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **174**(1), 175–194.

Banks, J., Marmot, M., Oldfield, Z. and Smith, J. P. (2006), 'Disease and disadvantage in the United States and in England', *JAMA: The Journal of the American Medical Association* **295**(17), 2037 –2045.

Böckerman, P. and Ilmakunnas, P. (2009), 'Unemployment and self-assessed health: evidence from panel data', *Health Economics* **18**(2), 161–179.

Cox, B., Huppert, F. and Whichelow, M., eds (1993), *The Health and Lifestyle Survey: Seven Years on: A Longitudinal Study of a Nationwide Sample, Measuring Changes in Physical and Mental Health, Attitudes and Lifestyle*, Dartmouth Pub. Co, Aldershot, Hants., England.

Cox, C., Chu, H., Schneider, M. F. and Muñoz, A. (2007), 'Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution', *Statistics in Medicine* **26**(23), 4352–4374.

Cutler, D., Deaton, A. and Lleras-Muney, A. (2006), 'The determinants of mortality', *The Journal of Economic Perspectives* **20**(3), 97–120.

Dalstra, J., Kunst, A., Borrell, C., Breeze, E., Cambois, E., Costa, G., Geurts, J., Lahelma, E., Van Oyen, H., Rasmussen, N., Regidor, E., Spadea, T. and Mackenbach, J. (2005), 'Socioeconomic differences in the prevalence of common chronic diseases: an overview of eight European countries', *International Journal of Epidemiology* **34**(2), 316 –326.

Deaton, A. (2002), 'Policy implications of the gradient of health and wealth', *Health Affairs* **21**(2), 13 –30.

Department of Health and Human Services (1989), *Reducing the health consequences of smoking. 25 years of progress*, US Government Printing Office, Washington, DC.

Department of Health and Human Services (2004), 'Smoking and Tobacco Use; 2004 Surgeon General's Report', http://www.cdc.gov/tobacco/data_statistics/sgr/2004/index.htm.

Department of Health and Human Services (2010), 'A Report of the Surgeon General: How Tobacco Smoke Causes Disease', http://www.surgeongeneral.gov/library/tobaccosmoke/report/index.html.

Doll, R. (1998), 'Uncovering the effects of smoking: historical perspective', *Statistical Methods in Medical Research* **7**(2), 87 –117.

Doll, R. and Hill, A. B. (1954), 'The mortality of doctors in relation to their smoking habits: a preliminary report', *BMJ* **328**(7455), 1529–1533.

Douglas, S. and Hariharan, G. (1994), 'The hazard of starting smoking: Estimates from a split population duration model', *Journal of Health Economics* **13**(2), 213–230.

Estève, J., Benhamou, E. and Raymond, L. (1994), 'Statistical methods in cancer research. volume IV. descriptive epidemiology.', *IARC scientific publications* (128), 1.

Gordo, L. R. (2006), 'Effects of short- and long-term unemployment on health satisfaction: evidence from German data', *Applied Economics* **38**(20), 2335.

Hiatt, R. A. and Breen, N. (2008), 'The social determinants of cancer: A challenge for transdisciplinary science', *American Journal of Preventive Medicine* **35**(2, Supplement 1), S141–S150.

Honoré, B. E. and Lleras-Muney, A. (2006), 'Bounds in competing risks models and the war on cancer', *Econometrica* **74**(6), 1675–1698.

Jones, A. M., Rice, N., Bago d'Uva, T. and Balia, S. (2007), *Applied Health Economics*, Routledge advanced texts in economics and finance, Routledge, London.

Link, B. G. and Phelan, J. (1995), 'Social conditions as fundamental causes of disease', *Journal of Health and Social Behavior* **35**, 80–94.

Marmot, M., Rose, G., Shipley, M. and Hamilton, P. J. (1978), 'Employment grade and coronary heart disease in British civil servants', *Journal of Epidemiology and Community Health* **32**(4), 244–249.

Marmot, M., Smith, G. D., Stansfeld, S., Patel, C., North, F., Head, J., White, I., Brunner, E. and Feeney, A. (1991), 'Health inequalities among British civil servants: the Whitehall II study', *Lancet* **337**(8754), 1387–1393.

Mathers, C. D. and Schofield, D. J. (1998), 'The health consequences of unemployment: the evidence.', *The Medical Journal of Australia* **168**(4), 178.

Peto, R., Lopez, A., Boreham, J. and Thun, M. (2006), 'Mortality from smoking in developed countries, 1950-2000'.

Singh, G. K., Miller, B. A., Hankey, B. F., Edwards, B. K., Singh, G. K., Miller, B. A., Hankey, B. F. and Edwards, B. K. (2003), 'Area socioeconomic variations in US cancer incidence, mortality, stage, treatment, and survival, 1975-1999'.

University of Cambridge Clinical School (2009), 'HALS deaths and cancer working manual, june 2009'.

US National Cancer Institute (2006), 'Cancer trends progress report - stage at diagnosis'.

Vallejo-Torres, L. and Morris, S. (2010), 'The contribution of smoking and obesity to income-related inequalities in health in England', *Social Science & Medicine* **71**(6).

Vineis, P., Alavanja, M., Buffler, P., Fontham, E., Franceschi, S., Gao, Y. T., Gupta, P. C., Hackshaw, A., Matos, E., Samet, J., Sitas, F., Smith, J., Stayner, L., Straif, K., Thun, M. J., Wichmann, H. E., Wu, A. H., Zaridze, D., Peto, R. and Doll, R. (2004), 'Tobacco and cancer: Recent epidemiological evidence', *Journal of the National Cancer Institute* **96**(2), 99 –106.

Wilkinson, R. and Pickett, K. (2010), *The Spirit Level: Why Equality Is Better for Everyone*, Penguin, London.