DET ØKONOMISKE RÅD
SEKRETARIATET

Mortality and Socio-economic Differences in a Competing Risks Model

Jakob Roland Munch and Michael Svarer

Working Paper 2001:1

The Secretariat publishes a series of Working Papers, which primarily contains detailed information on the analyses of selected subjects of the reports by the Chairmanship. The Secretariat is responsible for the series of Working Papers.

Peder Andersen
Head of Secretariat

ISSN 0907-2977 (Working Paper - Danish Economic Council)
List of previous Working Papers: see last page.

To order, contact:

Danish Economic Council
Secretariat
Adelgade 13, 5.
1304 København K

Phone: 33 13 51 28
Fax: 33 32 90 29
E-mail: dors@dors.dk
Home page: www.dors.dk

Symbols:
• Not applicable/data not available.
Due to rounding, the figures in tables may not add up to the totals.

Mortality and Socio-economic Differences in a Competing Risks Model

Jakob Roland Munch and Michael Svarer

Working Paper 2001:1

Abstract: This paper presents evidence on how mortality in Denmark is related to different socio-economic indicators. By use of a unique and extensive sample of the Danish population, we examine how mortality is related to factors such as education, occupation, skill level and income for the years 1992-97. We employ a competing risks proportional hazard model to allow for different causes of death. This turns out to be of importance as some factors have unequal (and sometimes opposite) influence on the cause-specific mortality rates. Particularly, we find that the often found inverse correlation between socio-economic status and mortality to a large degree is absent for women that die due to cancer.

Keywords: Mortality, competing risks duration model

JEL: C41, I12

Table of contents

1 Introduction

Since the 1970s life expectancy in Denmark has had a slower growth than in most other advanced countries. This is reflected by the fact that among the fifteen EU-countries only three countries had a higher life expectancy than Denmark in 1980, while in 1998 only one country had a lower life expectancy.[1] At the same time Denmark is among the most wealthy countries in EU. According to the 1998 OECD statistic on GDP per capita Denmark is the second richest country in EU, only succeeded by Luxembourg. Generally, life expectancy and GDP per capita have a very high positive correlation (see Sickles & Taubman (1997)). It is therefore remarkable that Denmark is at the lower end of the life expectancy distribution in EU. The explanation for this rank is presumably a complex mixture of supply side factors like individual behavior, life-style (e.g. attitudes toward smoking and alcohol use) and genetic factors in the population and demand side factors like access to treatment and quality of treatment. A thorough investigation of the Danish excess mortality problem therefore requires a very rich data set covering all parts of the mentioned factors.

Alongside the aggregate correlation between life expectancy and GDP per capita it is also found that high socio-economic status supports longevity (see e.g. Moore & Hayward (1990) and Martikainen (1995)). In the present study we focus on the more disaggregate level and investigate the association between socio-economic status and mortality at the individual level.

We use a very rich and extensive sample of the Danish population to examine how mortality is related to factors such as education, skill level, occupation, and income for the years 1992-97. The data set is formed by merging a number of administrative registers for a 10 percent sample of the Danish population. The registers do not contain information about health behaviors or risk factors such as consumption of cigarettes or alcohol. Instead there are

0) Results from an earlier slightly different version of this paper have appeared in Section IV.3 of Danish Economic Council (2000). Financial support from the Danish Social Science Research Council is gratefully acknowledged by Michael Svarer. We thank Niels Henning Bjørn, Mette Ejrnæs, Anne Kristine Høj and Michael Rosholm for useful comments and Jens Chr. Thellesen for excellent research assistance.

1) Life expectancy in Denmark in 1980 and 1998 were 74 and 76 years respectivly. The number for 1980 was only topped by Netherlands, Spain and Sweden, and in 1998 only Portugal had a lower life expectancy.

detailed information about marital status, education level, skill level, occupation, and income. That is, we try to measure life-style broadly and indirectly through several socio-economic variables.

The available panel data set allows us to perform estimations of a duration model, thus capturing the mortality experiences of different cohorts as they age. In doing so we take full advantage of the longitudinal nature of data. Furthermore, we employ a rather flexible competing risks proportional hazard model to allow for different causes of death. The results of this model is held up against results from a duration model ignoring different death causes, i.e. a single risk model. This clearly illustrates the relevance of a competing risks model, as some factors have unequal (and sometimes opposite) influence on the cause-specific hazard rates. The estimates from the single risk model mask these differences, and so important results may be overlooked if only the single risk model is estimated and analysed.

Our findings reveal that for men the negative correlation between socio-economic status and mortality generally prevails, but for women we find quite different results. For women in the age group 50-69 years old we find that factors such as being married, having a high income and having a long education are not significantly associated with higher life expectancy. It turns out that these surprising results primarily are due to the cancer mortality risk.

The next section describes in detail the data set and in section 3 the competing risks proportional hazard model is specified. Section 4 presents results of our estimations and finally section 5 gives a brief conclusion.

2 Data

Data used in this study are drawn from a large Danish panel data set made available by Statistics Denmark. The complete data set consists of a random sample of 10 % of the Danish adult population with information taken from several Danish administrative registers. The data set comprises information on a large number of demographic and socio-economic variables as well as information of death causes for all deaths in the years 1992-1997.[2] The extent

2) Information on the demographic and socio-economic variables for the year 1988 is also available.

of the data set makes it very powerful to examine the relationship between mortality and socio-economic characteristics.

We have divided the sample into four groups; according to gender and according to cohort. That is, we consider the two groups of men and women with the age of 40-49 years in 1992 and the two groups of men and women with the age of 50-69 years in 1992. Individuals below the age of 40 in 1992 is ignored because there are too few death events to identify effects from the covariates. In the other end of the age distribution we leave out individuals above the age of 69 in 1992, since in this cohort almost all individuals are retired and so there are no or misleading information on several socio-economic indicators. The observed lifetime is measured precisely as we have information on the date of birth for all individuals and the date of death for those who died in 1992-1997.

We explicitly take into account cause-specific mortality by classifying all deaths into a few important groups of death causes. The division is based on the eighth and tenth revision of the International Classification of Diseases (ICD). In the younger cohorts we have both estimated a single risk model and a competing risks model with two cause-specific hazard rates, where the two causes of death are cancer related illnesses and its complements. For the older cohorts we estimate an additional cause-specific hazard rate, namely that of death due to circulatory diseases. Table 1 shows the number of deaths in each of the four age-gender groups and their distribution on death causes. The figures show notable differences between men and women as cancer is a much more frequent cause of death for women than for men.

Below follows a brief description of the variables of the data set that are used in the subsequent analysis. Table 2 presents means for all explanatory variables for the four age-gender groups, and it is seen that there are more than 34.000 persons in each cohort. Note that for those persons who die during the period 1992-97 all covariates (except income) are observed the year prior to that of death. For those who survive the entire period, covariates are observed in the final year, 1997.

Among demographic indicators we consider three different indicator variables; *copenhagen* is an indicator variable that takes the value 1 if the person lived in the city of Copenhagen. Copenhagen is the capital and by far the largest city in Denmark. *married* is also an indicator variable, and is compared to other studies (e.g. Hu & Goldman (1990), Smith & Zick (1994), and Lillard & Panis (1996)) a somewhat crude measure, since we do not distinguish between never-married and divorced or widowed individuals. The presence of children is captured by the variable *kids*. It was only possible to control for the presence of children for persons in the younger cohort with sufficient precision, so the variable is left out of the analysis for the older cohorts.

Several different measures of socio-economic status are included in the analysis; education, skill level, occupation, home-ownership, and income. With respect to education there are five different categories defined in terms of the level of the most advanced completed education. The lowest level of education consists of persons with just primary schooling (No education). The next education category are those with vocational training and high school graduates (Vocational Education). Graduates have been grouped into the last three categories (Education 1 (shortest), 2 and 3 (longest)) based on the length of time needed to obtain a degree. Persons without education beyond primary school are used as reference group.

The next socio-economic factor is skill level, and again it is controlled for by constructing a number of indicator variables. The first category consists of self-employed persons which encompasses e.g. farmers and fishermen. There are three different levels of salaried workers; high, medium and low, and there are skilled and unskilled workers. For individuals who have left the labour market there are two types of retirement indicators namely *early-retirement* and *retirement*. Persons on the *early-retirement* program are individuals, who have had a long active labour market career, and who have retired at the age of 60 on favorable conditions. Thus only individuals in the older cohort had the opportunity to retire under these terms. As noted above we use information on the variable from the last year the persons are alive, and in this way several persons – particularly for the older cohorts – will be classified as retired. However, for most persons we have information about the latest skill level prior to retirement, and we have chosen to take this into account. That is, the variable for retirement measures those persons who have been retired through all the years 1988, 1992-1997. Skilled workers is the reference group below.

Another indicator of social status is occupation, and there are eight different categories; *agriculture/fishing, manufacturing, construction, trade, transport, public sector, service* and *unknown occupation*. Again, by using the occupation for the last year the persons are alive, most retired persons will be classified as having unknown occupation. Therefore we have chosen to assign the latest occupation different to unknown occupation. One might suspect the variables for skill level and occupation of being highly correlated as for instance self employed persons typically might work in the agriculture/fishing sector. However, we have calculated correlation coefficients, and they reveal no strong correlations between any of the socio-economic indicators. Those working in the service sector is the reference group.

Finally, two economic indicators for home ownership and income are included. It is seen from Table 2 that men are more often home owners than women, reflecting the fact that for married couples it is typically husbands who are registered as the owner. The income variable is gross income in the year 1988. We have chosen income in this year so as to ensure at least 4 years between income is observed and the date of death for any person in the sample. Thereby we avoid the tendency for income to fall due to reduced labour market activity in the years preceding death.

3 Methodology

In this section we describe the statistical method applied in the study. The mortality rate is given by the hazard function which is the probability of dying in the interval $(t; t + dt]$ given survival until t. Death can occur due to the various causes described above, and we allow for different causes of death by specifying cause-specific hazard functions. The vector of explanatory variables, x, is included to capture the influence of the various exogenous socio-economic factors on the mortality rate. Let the continuous stochastic variable T, $T \in (0, \infty)$, denote age, and let the discrete stochastic variable i, $i = 1, 2, ..., I$, denote cause of death (henceforth denoted destinations). The cause-specific hazard rate is then given by

$$h_i(t|x) = \lim_{dt \to 0} \frac{\mathbf{P}(t < T \leq t + dt, I = i | T > t, x)}{dt}. \tag{1}$$

Given that the destinations are mutually exclusive and exhaust all possible destinations, the marginal hazard function is the sum of the cause-specific hazard functions:

$$h(t|x) = \sum_{i=1}^{I} h_i(t|x). \tag{2}$$

Each cause-specific hazard functions is specified as a proportional hazard model. That is, the hazard is the product of the baseline hazard, which captures the time dependence and a function of observed characteristics, x,

$$h_i(t|x) = \lambda_i(t) \cdot \varphi_i(x), \tag{3}$$

where $\lambda_i(t)$ is the baseline hazard and $\varphi_i(x_t)$ is the scaling function specified as $\exp(x\beta)$.

Normally, the contribution to the likelihood function for a given individual is simply the density[3], $f(t)$. In this case, however, we have to adjust for the fact that we only observe individuals who have survived until 1992. Let r denote the age of individual j in 1991. The contribution for this individual to the likelihood function is then

$$\begin{aligned}
f(t_j|t_j > r) &= \frac{f(t_j)}{P(t_j > r)} \\
&= \frac{h(t_j)S(t_j)}{S(r)} \\
&= \frac{h(t_j)\exp\left(-\int_0^{t_j} h(s)ds\right)}{\exp\left(-\int_0^r h(s)ds\right)} \\
&= h(t_j)\exp\left(-\int_0^{t_j} h(s)ds + \int_0^r h(s)ds\right) \\
&= h(t_j)\exp\left(-\int_r^{t_j} h(s)ds\right),
\end{aligned}$$

where $S(r) = \exp\left(-\int_r^{t_j} h(s)ds\right)$ is the survivor function.

3) Suppressing dependency on covariates.

3.1 The likelihood function

Since the hazard, by nature, is increasing with age (senescence) we specify the baseline hazard to accommodate positive duration dependence.[4] Specifically, we specify a piecewise constant baseline hazard with splitting times $\tau_0 = a, \tau_1, ..., \tau_K = +\infty$, where a is the age of the youngest person in the sample. Notice, that the baseline can attain arbitrary flexibility by increasing the number of intervals. The value of the baseline hazard in the k'th interval is given by λ_i^k. Let $k(t) : \Re_+ \curvearrowright \{1, 2, ..., K - 1, K\}$ be a function that maps the duration, t, into interval k. Based on the different causes of death we define I destinations indicators $d_i = \mathbb{I}_{\{I=i\}}$, where $\mathbb{I}_{\{\cdot\}}$ is an indicator function for the event in the brackets. The contribution for an individual with observed characteristics, x, is then

$$\begin{aligned} \mathcal{L}(\theta) &= h_i\left(t|x_i\right)^{\sum_{i=1}^I d_i} \exp\left[-\sum_{i=1}^{I}\int_r^t h_i\left(s|x_i\right) ds\right] \\ &= \left(\lambda_i^{k(t)} \cdot \exp\left[x_i\beta_i\right]\right)^{\sum_{i=1}^I d_i} \qquad (4) \\ &\quad \times \exp\left[-\sum_{i=1}^{I}\left\{\exp\left[x_i\beta_i\right] \cdot \left(\begin{array}{c} \sum_{h=r+1}^{k(t)-1} \lambda_i^h\left(\tau_h - \tau_{h-1}\right) \\ +\lambda_i^{k(t)}\left(t - \tau_{k(t)-1}\right) \end{array}\right)\right\}\right]. \end{aligned}$$

Note, that uncompleted (i.e. right censored) observations only contribute with the survivor function as $\sum_{i=1}^I d_i = 0$.

4) Our choice of model is somewhat guided by the results found by Behrman, Sickles & Taubman (1990). They estimate mortality rates on a sample of U.S. veterans, and conduct several sensitivity analyses concerning the statistical specification of there model. Briefly summarised, they find that their results were sensitive to the sample length due to the amount of right censored observations, but not sensitive to parametric specification of the hazard function or to the inclusion of terms correcting for unobserved heterogeneity (frailty). Behrman, Sickles, Taubman & Yazbeck (1991) estimate mortality rates on US retirement data and also find that their results are insensitive to the inclusion of frailty. In the following we employ a rather flexible hazard function that is computational attractive and we do not correct for unobserved heterogeneity.

3.2 Remaining life expectancy (RLE)

The interpretation of estimation results in hazard models are more complicated than in other statistical models, since the size of the estimated coefficient for the explanatory variables is not straightforward to evaluate. In order to accommodate this feature of hazard models we quantify the effect of the personal characteristics on the mortality rate by calculating the expected duration of remaining life. That is, for given personal characteristics it is possible to calculate how the expected remaining life time changes when the size of a covariate changes. Below we illustrate how the calculations are done.

Suppressing the dependency of the hazard function on x, the integrated hazard is

$$\begin{aligned} H(t) &= \int_0^t h^{k(s)} ds \\ &= H^{k(t)-1} + h^{k(t)}(t - \tau_{k(t)-1}), \end{aligned}$$

where $H^{k(t)-1}$ is the integrated hazard up to the beginning of the k'th interval.

The expected duration is then (see Lancaster (1990))

$$\begin{aligned} E[T] &= r + \int_r^\infty S(t)dt \\ &= r + \int_r^\infty e^{-H(t)} dt \\ &= r + \sum_{k=r+1}^{K} \int_{\tau_{k-1}}^{\tau_k} e^{-H^{k-1} - h^k(t - \tau_{k-1})} dt \\ &= r + \sum_{k=r+1}^{K} \left(-\frac{1}{h^k} e^{-H^{k-1} - h^k(\tau_k - \tau_{k-1})} + \frac{1}{h^k} e^{-H^{k-1} - h^k(\tau_{k-1} - \tau_{k-1})} \right) \\ &= r + \sum_{k=r+1}^{K} \frac{1}{h^k} \left(e^{-H^{k-1}} - e^{-H^k} \right). \end{aligned}$$

This can also be expressed as

$$E[T] = r + \sum_{k=r+1}^{K} \frac{1}{h^k} P(\tau_{k-1} < T \leq \tau_k). \quad (5)$$

That is, the RLE is simply the sum of the inverse interval-specific hazard functions weighted with the interval-specific occurrence probability.

4 Results

“I would rather discover a single causal connection than win the throne of Persia”

Democritus (460-370 BC)

The famous quote by the Greek philosopher Democritus very precisely describes a prevailing concern in mortality studies (as well as in a lot of other disciplines), namely that of cause and effect. In this study we present correlations between mortality rates and different socio-economic covariates. By doing so we are able to identify socio-economic differences in mortality rates. However, we are not able to identify whether specific socio-economic covariates affect the mortality rate through indirect effects that are derived from the fact that the reason why an individual carries certain characteristics is due to her underlying mortality risk. We have these limitations in mind when interpreting the results that follow.

Before the estimation results are discussed, it should be noted that subgroup differences in mortality experiences of the population were assessed by sequentially introducing the variables for residence in Copenhagen, marital status, education, skill level, occupation, home ownership, and income. Generally, we find that an additional variable reduces the effect of previously entered variables as in Moore & Hayward (1990). For instance for the older cohort of women the effects of vocational education and education level 1 are each time reduced somewhat by the sequential addition of variables controlling for skill level, occupation, home owners and income. This emphasizes the importance of using several indicators simultaneously in order to understand the extent to which socio-economic status affects mortality.

Only very few variables change from having a significant effect on mortality to having an insignificant effect. For the older cohort of women the effect of being retired turns insignificant by introduction of the variables for occupation, which most likely is due to the correlation between the variables *unknown* occupation and *retired*. Also, the negative effect from having education level 3 becomes insignificant by entering skill level into the estimation. For this reason we have chosen not to present the result for models other than the full model that includes all demographic and socio-economic indicators.

In the following subsections we discuss the results presented in Table 3-7.[5] First, we go through the effects of the various demographic and socio-economic factors on the cause-specific mortality rates. Then, in order to quantify the effects from the covariates, we calculate remaining life expectancies in the last section.

4.1 Marriage and kids

A substantial amount of literature have found that married individuals have significantly lower risks of mortality than their single counterparts (see e.g. Smith & Zick (1994) for a survey), and that these results are particularly strong for men, see Lillard & Panis (1996). Whether this difference is due to selection mechanisms of healthier individuals into marriage or due to causal processes, typically known as marriage protection (where marriage induces are more healthy life style), has been on the research agenda for long (see Goldman (1993) for a discussion). Yet, there is a potentially counter-acting effect. Lillard & Panis (1996) argue that if marriage has a positive effect on longevity, then individuals in poor health or with higher mortality risk have a greater incentive to marry and gain that protection. They test their hypothesis on a sample of American men in a simultaneous model of marital status, health and mortality, and when they correct for unmeasured characteristics they find that healthier men are less likely to marry. However, based on the unmeasured characteristics they also find that men who marry earlier tend to be healthier than men who remain unmarried. That is, their results support both a positive and an adverse selection effect. In their application the positive selection effect dominates if they do not take unmeasured factors into account.

5) We have excluded the estimated baseline hazards for expositional reasons. They, however, exhibits positive duration dependence as expected, and are available upon request.

Regardless of which factor, if not all, drives the results we do find that for men marriage is associated with lower mortality risk[6] (see Table 4 and 6). However, for the cancer sub-hazard the effect of marriage is insignificant. Our findings for women are particularly remarkable, though, as the cause-specific hazard rates reveal a significantly *positive* effect of marriage on cancer mortality and a significantly negative effect of marriage on the risk of dying of other diseases (see Table 3 and 5). For the younger cohort of women an overall negative effect of marriage on mortality remain, but for the older cohort the overall effect is insignificant. For this cohort there is no significant effect of marriage on the circulatory disease sub-hazard. These findings would have been blurred had we not applied a competing risks duration model. In the terminology of Lillard & Panis (1996) it appears that the adverse selection effect dominates for some sub-samples of women.[7] Ellwood & Kane (1989) find, in a sample of American women and men over 65 years old, that married women have a higher mortality risk than single women. They argue that the result could be caused by the traditional roles of husband and wife: married men live longer because they are cared for by their wives. In the process, wives' health may be endangered. In a Danish context this effect could be very important, since Danish women have very high labour market participation rates.

It should be noted that for the younger cohort we have included a variable capturing whether the individuals have children. For women, we find a significant negative effect on the mortality risk, even in the cancer sub-hazard. The same pattern emerges for men, although the effect in the cancer sub-hazard is insignificant.

6) Note, that the reference group is not married individuals and that this group therefore contains, divorced, separated, never married and widowed individuals. Potentially some of the sub-groups could have lower mortality risk than the married individuals and in that respect distort our findings. Previous research reveal mixed results when looking at subgroups. Whereas Hu & Goldman, (1990) and Lillard & Panis (1996) find that a higher mortality risk is also found for the sub-groups, Behrman et al. (1991) find for men that married and divorced/separated have lower mortality risk than their counterparts.

7) We could, of course, investigate this feature in more detail by endogenizing marital status in the mortality hazard. We leave this exercise for future research.

4.2 Education and income

Both education and income are from a theoretical perspective negatively associated with mortality risk. First, it is generally perceived that more educated individuals make better decisions and can process new information more effectively which consequently should lead to more healthy life-styles. Second, increases in income permit individuals to devote more time to health production which reduces sick time and subsequently mortality risk.

On an empirical level these predictions have been found to hold more or less consistently in a number of studies (see Sickles & Taubman (1997)) for a survey and see also e.g. Hayward, Grady, Hardy & Sommers (1989), Moore & Hayward (1990), Smith & Zick (1994) and Hurd, McFadden & Merril (1999). In this study, we find for men that the negative relationship between education/income and mortality risk prevails, cf. Table 4 and 6.[8] For women in the older cohort, however, we find quite different results. Here, income has no significant effect on mortality, and the effect of education on mortality is non-monotone. Table 5 column 1 shows that for relatively low educated women (vocational education and education group 1) there is a negative effect on the mortality compared to women with no education, but this effect disappears for women with medium and high education. The cause-specific hazard rates reveal that the negative effects are driven by the sub-hazard for circulatory diseases. For women in education group 2, there is a significant positive effect in the cancer hazard. A somewhat similar result is found by Martikainen (1995). He finds that for Finish women there is a positive association between education and death due to breast cancer; a result that doesn't carry over to other forms of cancer. Also, Goldman, Takahashi & Hu (1995) find related results. When investigating deaths among Japanese single women they find that death due to cancer is positively related to income.

The indicator variable capturing whether the individual is a home owner is, like education and income, supposed to capture socioeconomic status. Generally, we find the expected significant negative effect on mortality risk, but it is also clear that the result is driven by a negative effect on the risk of dying of other diseases than cancer and circulatory diseases.

8) The reference category for education is no education beyond elementary school.

4.3 Skill level

There are, compared to skilled workers, very few significant differences with respect to mortality risk.[9] For women, we only find significantly lower mortality risks in the sub-hazard for deaths due to other diseases. For men in the older cohort, however, there is an interesting finding. Men on *early-retirement* have a substantially lower mortality risk. The *early-retirement* program was introduced in 1979, and it gives individuals, who have had an active labour market career, the choice to retire at the age of 60. The program was introduced to provide older workers with the opportunity to stop working before their physical health deteriorate. As such, the intention of the program, at least as formulated by the politicians, was to relieve worn-out workers from the burden of hard physical work tasks. In this respect, we would expect individuals on *early-retirement* to have a relatively higher mortality risk due to bad health. Our results, if anything, reveal the opposite finding, which suggest that either it is not only worn-out individuals that take advantage of *early-retirement,* or that individuals, who leave the labor force actually experience improved health. This effect could, however, be caused by the composition of the individuals eligible for *early-retirement,* since a long and active labour market career[10] is stipulated for entering the programme. Several authors have found a positive association between active labour market participation and longevity (see next sub-section).

4.4 Occupation

As argued by Sickles & Taubman (1997) occupation is perhaps the most clear example of a potential reverse causality between covariates and health, due to the reliance of certain occupations on the level of physical activity. On the other hand, occupation is a crucial covariate and according to Moore & Hayward (1990) the most accurate single indicator of social status and life-style. On top of that, occupation also defines the environments in which persons are exposed to hazardous working conditions, pollution, and emotional stress.

9) The reference category is skilled labour.

10) In fact, individuals should be active on the labour market in 20 out of the latest 25 years in order to be eligible for early-retirement.

For the younger cohort we only find significant effects from individuals with unknown occupation.[11] This group of individuals, which primarily consists of individuals outside the labour force, has a higher risk of mortality.[12] The result is in accordance with other studies; Jones & Goldblatt (1987) find that women outside the labour market have higher mortality rates than women in the labour market, whereas Sorlie & Rogot (1990) find similar results for men. For the older cohort the positive effect is sustained, and here it is driven by all cause-specific hazard rates (except death due to circulatory diseases for women).

The other significant results for this cohort are presumably closely connected to work environments. Men, who works in the agricultural sector have significantly lower mortality risk, whereas men working in the manufacturing, construction or transport sector have higher mortality risk. While the latter effects only pertain in the cancer sub-hazard, the former effect is found in the two other sub-hazard rates. For women, we also find that those working in the manufacturing sector have higher mortality risk due to cancer. The negative effect found in the circulatory hazard for public employed female workers could be attributed to less stressful working environments.

4.5 Remaining life expectancy (RLE)

To facilitate comparisons between persons with different socio-economic characteristics we have calculated remaining life expectancies as described in section 3.2. Before discussing the findings some comments on how the calculations are done are in order.

We use expression (5) and the estimation results from the single risk model for the older cohorts to calculate RLE. Since we only estimate mortality rates for individuals up to 69 years old we have to extrapolate the baseline hazard for ages above 69. Here, we assume that the mortality risk stays constant. This is clearly a misspecification which implies that the RLE's are upwards biased. The RLE's are calculated based on a constructed person profile. That is, we choose specific characteristic and calculate the RLE for this hypothetically constructed individual. All comparisons are therefore based on this

11) The reference category is service sector.

12) Note that this effect is insignificant for the cancer hazard.

constructed individual and doesn't necessarily carry over to other person profiles. Still, the exercise can be fruitful bearing its limitations in mind. The standard persons we have constructed[13] have RLE's equal to 30.6 years for women and 27.4 years for men.[14]

In Table 7 we present different simulations. The positive effect on the mortality risk from living in the Copenhagen metropolitan area is quite substantial. Women (men) can expect to live 2.3 (1.3) years less than their provincial counterparts. It is also interesting to note that factors like marriage, high education and high income that is typically connected with high socio-economic status have significant positive effect on the RLE for men, but insignificant effect for women. On top of that the effects for men is rather large; not being married takes 2.2 years from RLE and being in the highest education category give 3.1 years in RLE. Raising income far beyond average income by doubling it only gives .5 years in RLE, even though the effect is highly significant in Table 6.

5 Conclusion

In this article we investigate the relationship between mortality and different demographic and socio-economic indicators taking into account the distinction between different causes of death. The importance of this distinction was examined by comparing the results from a single risk model with those from a competing risks model where death may occur due to cancer, circulatory diseases and other diseases. Estimation of the competing risks model shows that the covariates have different effects on the cause-specific mortality rates with respect to significance as well as signs.

We find that the expected negative association between socio-economic status and mortality indeed prevails for men, but to a much smaller extent for women. The somewhat surprising results for women are especially pronounced in the cancer hazard, and for future research it would be interest-

13) Standard person: Aged 50 in 1992, living outside the Copenhagen metropolitan area, married, no education, unskilled worker, rented housing, working in the service sector, income of 200,000 DKr.

14) According to Statistics Denmark (2000) the RLE for a woman (man) aged 50 in 1992 was 29.9 (25.7) years.

ing to investigate further whether the result is due to specific cancer illnesses such as breast cancer.

To conclude that the absence of an inverse relationship between socio-economic status and mortality that is found for women is the sole explanation for the excess mortality in Denmark is of course out of order. Still, the findings suggest that possible explanations can be found. To draw more firm conclusions a more elaborate data set including information on life-style and demand side factors is needed, and we encourage the collection of such data sets.

References

Behrman, J. R., Sickles, R. C. & Taubman, P. (1990), 'Age-specific death rates with tobacco smoking and occupational activity: Sensitivity to sample length, functional form, and unobserved frailty', *Demography* **27**, 267–284.

Behrman, J. R., Sickles, R. C., Taubman, P. & Yazbeck, A. (1991), 'Black-white mortality inequalities', *Journal of Econometrics* **50**, 183–203.

Danish Economic Council (2000), *Danish Economy, Spring 2000*, Copenhagen.

Ellwood, D. T. & Kane, T. J. (1989), The american way of aging: An event history analysis. NBER Working Paper No. 2892.

Goldman, N. (1993), 'Marriage selection and mortality patterns: Inferences and fallacies', *Demography* **30**, 189–195.

Goldman, N., Takahashi, S. & Hu, Y. (1995), 'Mortality among japanese singles: A re-investigation', *Population Studies* **49**, 227–239.

Hayward, M. D., Grady, W. R., Hardy, M. A. & Sommers, D. (1989), 'Occupational influences on retirement, disability, and death', *Demography* **26**, 393–409.

Hu, Y. & Goldman, N. (1990), 'Mortality difference by marital status: An international comparison', *Demography* **27**, 233–248.

Hurd, M. D., McFadden, D. & Merril, A. (1999), Predictors of mortality among the elderly. NBER Working Paper No. 7440.

Jones, D. R. & Goldblatt, P. O. (1987), 'Cause of death in Widow(Er)S and spouses', *Journal of Biosocial Sciences* **19**, 107–121.

Lancaster, T. (1990), *The Econometric Analysis of Transition Data*, Cambridge University Press, New York.

Lillard, L. A. & Panis, C. W. A. (1996), 'Marital status and mortality: The role of health', *Demography* **33**, 313–327.

Martikainen, P. (1995), 'Mortality and socio-economic status among finnish women', *Population Studies* **49**, 71–90.

Moore, D. E. & Hayward, M. D. (1990), 'Occupational careers and mortality of elderly men', *Demography* **27**, 31–53.

Sickles, R. & Taubman, P. (1997), Mortality and morbidity among adults and the elderly, *in* M. R. Rosenzweig & O. Stark, eds, 'Handbook of Population and Family Economics', Vol. 1A, Elsevier Science, Amsterdam, chapter 11, pp. 559–643.

Smith, K. R. & Zick, C. D. (1994), 'Linked lives, dependent demise? survival analysis of husbands and wives', *Demography* **31**, 81–93.

Sorlie, P. & Rogot, E. (1990), 'Mortality by employment status in the national longitudinal mortality study', *American Journal of Epidemiology* **132**, 983–992.

Statistics Denmark (2000), *Statistisk Tiårsoversig 2000*, Copenhagen.

A Tables

Table A.1 Causes of deaths

	Women		Men	
Age, year	40-49	50-69	40-49	50-69
Number of deaths (1992-1997)	548	3119	794	4445
Number of individuals	34134	48829	35074	45962
Cause of death	Percentage			
Cancer	50.5	44.1	28.0	33.5
Circulatory diseases	11.5	23.9	20.9	35.1
Other	38.0	32.0	51.1	31.3
All	100.0	100.0	100.0	99.9

Table A.2 Means for explanatory variables

	Women		Men	
Age, year	40-49	50-69	40-49	50-69
Copenhagen	0.082	0.090	0.082	0.080
Married	0.722	0.592	0.708	0.731
Kids	0.897		0.815	
Vocational education	0.380	0.252	0.447	0.366
Education 1 (shortest)	0.082	0.030	0.060	0.037
Education 2	0.111	0.066	0.102	0.066
Education 3 (longest)	0.035	0.013	0.074	0.048
No education	0.392	0.639	0.317	0.483
Self employed	0.045	0.039	0.118	0.145
Salaried worker				
- high level	0.056	0.015	0.161	0.097
- medium level	0.156	0.059	0.146	0.079
- low level	0.302	0.114	0.097	0.055
Skilled	0.044	0.050	0.192	0.090
Unskilled	0.183	0.127	0.157	0.113
Early retirement	-	0.232	-	0.255
Retired	0.030	0.055	0.016	0.027
Other	0.184	0.309	0.113	0.139
Agriculture/fishing	0.017	0.035	0.036	0.077
Manufacturing	0.115	0.096	0.228	0.234
Construction	0.014	0.012	0.104	0.086
Trade	0.094	0.088	0.112	0.102
Transport	0.032	0.023	0.094	0.092
Public	0.532	0.409	0.254	0.211
Service	0.120	0.062	0.125	0.087
Unknown	0.052	0.224	0.030	0.071
Home owner	0.371	0.269	0.688	0.630
Income (in 1.000 DKr)	143	108	243	214
Number of persons	34134	48829	35074	45962

Table A.3 Estimation results for women, 40-49

	All causes		Cancer		Other causes	
Variables	Coeff.	Std.	Coeff.	Std.	Coeff.	Std.
Copenhagen	0.337	*0.134*	0.132	*0.215*	0.471	*0.174*
Married	-0.359	*0.099*	0.319	*0.159*	-1.006	*0.139*
Kids	-0.512	*0.117*	-0.652	*0.183*	-0.398	*0.162*
Vocational edu.	-0.119	*0.107*	-0.206	*0.157*	-0.033	*0.154*
Edu. 1 (shortest)	-0.620	*0.242*	-0.687	*0.336*	-0.494	*0.371*
Edu. 2	-0.402	*0.229*	-0.397	*0.299*	-0.428	*0.392*
Edu. 3 (longest)	-0.531	*0.362*	-0.778	*0.520*	-0.246	*0.515*
Self employed	0.846	*0.750*	0.576	*1.078*	0.428	*1.093*
Salaried worker						
- high level	0.490	*0.774*	0.317	*1.088*	-0.094	*1.155*
- medium level	0.623	*0.745*	0.407	*1.058*	0.109	*1.073*
- low level	0.651	*0.724*	0.397	*1.035*	0.271	*1.043*
Unskilled	0.690	*0.728*	0.045	*1.043*	0.799	*1.042*
Retired	0.840	*0.742*	0.240	*1.069*	0.892	*1.054*
Other	0.829	*0.725*	0.253	*1.046*	0.847	*1.030*
Agriculture/fishing	0.210	*0.330*	0.005	*0.496*	0.546	*0.459*
Manufacturing	0.022	*0.189*	0.289	*0.243*	-0.230	*0.324*
Construction	0.118	*0.404*	-1.010	*0.756*	0.773	*0.500*
Trade	0.032	*0.197*	0.155	*0.259*	-0.006	*0.325*
Transport	0.103	*0.288*	0.256	*0.361*	-0.246	*0.515*
Public	0.150	*0.150*	0.097	*0.210*	0.313	*0.228*
Unknown	0.899	*0.186*	0.360	*0.323*	1.160	*0.249*
Home owner	-0.313	*0.109*	-0.108	*0.148*	-0.528	*0.178*
Income	-0.063	*0.074*	-0.008	*0.109*	-0.134	*0.096*

Table A.4 Estimation results for men, 40-49

	All causes		Cancer		Other causes	
Variables	Coeff.	Std.	Coeff.	Std.	Coeff.	Std.
Copenhagen	0.131	*0.105*	-0.033	*0.241*	0.166	*0.119*
Married	-0.453	*0.094*	-0.149	*0.193*	-0.581	*0.110*
Kids	-0.391	*0.090*	-0.294	*0.191*	-0.436	*0.104*
Vocational edu.	-0.094	*0.082*	-0.281	*0.170*	-0.047	*0.097*
Edu. 1 (shortest)	-0.211	*0.189*	-0.449	*0.396*	-0.097	*0.223*
Edu. 2	-0.518	*0.196*	-0.562	*0.355*	-0.548	*0.246*
Edu. 3 (longest)	-0.374	*0.221*	-1.165	*0.584*	-0.182	*0.246*
Self employed	0.119	*0.164*	-0.028	*0.308*	0.146	*0.203*
Salaried worker						
- high level	-0.249	*0.194*	-0.829	*0.409*	-0.014	*0.230*
- medium level	-0.134	*0.168*	-0.220	*0.274*	-0.176	*0.220*
- low level	0.056	*0.167*	-0.391	*0.314*	0.224	*0.203*
Unskilled	0.127	*0.139*	-0.205	*0.253*	0.219	*0.171*
Retired	-0.191	*0.237*	-1.235	*0.584*	0.057	*0.267*
Other	0.518	*0.123*	-0.068	*0.275*	0.696	*0.157*
Agriculture/fishing	-0.499	*0.249*	-0.967	*0.518*	-0.282	*0.290*
Manufacturing	-0.163	*0.135*	-0.402	*0.266*	-0.068	*0.160*
Construction	-0.062	*0.151*	-0.286	*0.295*	0.047	*0.179*
Trade	-0.243	*0.160*	-0.071	*0.281*	-0.321	*0.200*
Transport	-0.111	*0.153*	-0.025	*0.286*	-0.130	*0.185*
Public	-0.197	*0.131*	-0.124	*0.257*	-0.217	*0.156*
Unknown	0.386	*0.160*	0.123	*0.372*	0.471	*0.184*
Home owner	-0.467	*0.087*	-0.260	*0.177*	-0.564	*0.105*
Income	-0.132	*0.037*	-0.168	*0.089*	-0.118	*0.043*

Table A.5 Estimation results for women, 50-69

	All causes		Cancer		Circulatory diseases		Other causes	
Variables	Coeff.	Std.	Coeff.	Std.	Coeff.	Std.	Coeff.	Std.
Copenhagen	0.329	*0.055*	0.361	*0.084*	0.156	*0.121*	0.419	*0.099*
Married	-0.019	*0.039*	0.247	*0.061*	-0.036	*0.081*	-0.342	*0.073*
Vocational edu.	-0.092	*0.048*	-0.015	*0.070*	-0.366	*0.113*	-0.010	*0.089*
Edu. 1 (short)	-0.304	*0.138*	-0.219	*0.196*	-1.144	*0.466*	0.076	*0.249*
Edu. 2	-0.034	*0.098*	0.249	*0.125*	-0.287	*0.233*	-0.405	*0.250*
Edu. 3 (long)	-0.290	*0.216*	0.018	*0.265*	-0.198	*0.529*	-1.040	*0.720*
Self employed	0.085	*0.338*	0.490	*0.524*	-0.307	*1.051*	-1.105	*0.525*
Salaried workers								
- high level	-0.277	*0.400*	0.328	*0.576*	-1.749	*1.450*	-2.288	*0.868*
- medium level	0.028	*0.342*	0.261	*0.527*	0.018	*1.057*	-1.580	*0.556*
- low level	0.089	*0.330*	0.542	*0.514*	-0.188	*1.043*	-1.414	*0.502*
Unskilled	0.163	*0.329*	0.453	*0.512*	-0.083	*1.039*	-1.070	*0.496*
Early retirement	-0.465	*0.328*	-0.063	*0.512*	-0.734	*1.037*	-1.871	*0.495*
Retired	0.348	*0.333*	0.188	*0.524*	0.089	*1.044*	-0.605	*0.503*
Other	-0.099	*0.328*	0.187	*0.511*	-0.582	*1.040*	-1.217	*0.493*
Agric./fishing	-0.268	*0.129*	-0.217	*0.195*	-0.041	*0.238*	-0.613	*0.288*
Manufacturing	0.206	*0.084*	0.353	*0.123*	0.014	*0.174*	0.141	*0.168*
Construction	-0.002	*0.195*	-0.166	*0.317*	-0.454	*0.517*	0.613	*0.298*
Trade	-0.082	*0.092*	0.163	*0.130*	-0.172	*0.189*	-0.320	*0.199*
Transport	0.068	*0.140*	0.346	*0.197*	0.111	*0.287*	-0.094	*0.305*
Public	-0.067	*0.070*	0.078	*0.104*	-0.362	*0.144*	0.008	*0.138*
Unknown	0.223	*0.073*	0.271	*0.113*	0.114	*0.146*	0.263	*0.136*
Home owner	-0.160	*0.047*	-0.035	*0.069*	-0.101	*0.098*	-0.362	*0.094*
Income	-0.005	*0.031*	0.044	*0.041*	-0.143	*0.084*	-0.086	*0.056*

Table A.6 Estimation results for men, 50-69

	All causes		Cancer		Circulatory diseases		Other causes	
Variables	Coeff.	Std.	Coeff.	Std.	Coeff.	Std.	Coeff.	Std.
Copenhagen	0.159	*0.051*	0.223	*0.091*	-0.004	*0.092*	0.283	*0.086*
Married	-0.268	*0.034*	-0.051	*0.062*	-0.104	*0.057*	-0.574	*0.064*
Vocational edu.	-0.006	*0.035*	0.080	*0.061*	-0.079	*0.060*	-0.016	*0.068*
Edu. 1 (shortest)	0.001	*0.088*	0.254	*0.136*	-0.261	*0.168*	-0.072	*0.188*
Edu. 2	-0.321	*0.085*	-0.262	*0.141*	-0.459	*0.150*	-0.263	*0.174*
Edu. 3 (longest)	-0.438	*0.098*	-0.303	*0.171*	-0.552	*0.169*	-0.648	*0.232*
Self employed	0.025	*0.082*	-0.134	*0.142*	0.094	*0.136*	0.137	*0.164*
Salaried worker								
- high level	-0.260	*0.099*	-0.177	*0.160*	-0.254	*0.169*	-0.409	*0.215*
- medium level	-0.084	*0.092*	0.106	*0.147*	-0.179	*0.160*	-0.299	*0.199*
- low level	-0.064	*0.097*	-0.119	*0.167*	-0.046	*0.162*	-0.006	*0.193*
Unskilled	0.007	*0.083*	0.041	*0.141*	-0.040	*0.140*	0.064	*0.163*
Early retirement	-0.308	*0.073*	-0.236	*0.123*	-0.375	*0.124*	-0.304	*0.147*
Retired	0.185	*0.103*	-0.048	*0.200*	0.052	*0.176*	0.484	*0.186*
Other	0.063	*0.077*	0.030	*0.132*	-0.028	*0.131*	0.253	*0.150*
Agriculture/Fishing	-0.434	*0.080*	-0.162	*0.138*	-0.659	*0.134*	-0.418	*0.165*
Manufacturing	0.139	*0.057*	0.250	*0.099*	0.069	*0.094*	0.216	*0.116*
Construction	0.098	*0.071*	0.262	*0.121*	0.118	*0.115*	0.061	*0.146*
Trade	0.022	*0.068*	0.101	*0.119*	-0.001	*0.112*	0.056	*0.137*
Transport	0.063	*0.069*	0.386	*0.114*	0.002	*0.113*	-0.142	*0.147*
Public	0.043	*0.060*	0.127	*0.104*	0.022	*0.099*	0.091	*0.122*
Unknown	0.430	*0.072*	0.390	*0.136*	0.405	*0.120*	0.552	*0.135*
Home owner	-0.106	*0.034*	0.038	*0.060*	-0.053	*0.056*	-0.416	*0.067*
Income	-0.035	*0.002*	-0.020	*0.017*	-0.040	*0.018*	-0.038	*0.023*

Table A.7 Remaining life expectancy

	Women	Men
Standard person	30.6	27.4
+ Copenhagen	28.3	26.1
- Married	30.5*	25.2
+ Edu. 1 (shortest)	32.4	27.4*
+ Edu. 3 (longest)	32.4*	30.5
+ House owner	31.6	28.2
+ Agriculture/fishing	31.9	30.8
+ Income 400,000 Dkr	30.7*	27.9
+ Copenhagen + edu. 3 + high level salaried worker + house owner + income 400,000 Dkr.	33.8*	32.2

* Not significant at the 5% level of significance.