Appendix to "Bounds in Competing Risks Models and the War on Cancer"

Bo E. Honoré Adriana Lleras–Muney

June 7, 2006

This appendix presents additional information that is referred to in "Competing Risks and the War on Cancer". The information is presented following the order of the paper.

1 Data

1.1 Hazard rates by demographic group, cause of death and year

Figures 1 and 2 show the hazard rates for white males, white females, black males and black females, for cancer and CVD separately. These hazard rates present in more detail the same trends that the summary statistics show. Hazard rates from CVD declined quite significantly in every decade for all groups. On the other hand, there is no discernible trend in cancer hazard rates. It is also clear that hazard rates are fairly different across demographic groups. From these graphs we also note that hazard rates are much more volatile among blacks, especially at older ages. This is true for both causes of death, but it is more pronounced for cancer rates. Censoring at age 80 alleviates the problem somewhat since hazard rates become even more volatile for older ages (not shown).

1.2 Some Data Issues

1.2.1 Age misreporting

In the census there is evidence of age heaping: individuals ages 50 and above tend to overstate their ages by "rounding up," which results in an unusually large population for ages ending in either 5 or 0. In our data age heaping is mostly an issue for blacks. Another important issue (that cannot be fully separated from age misreporting) is that the census undercounts certain groups of the population, especially blacks, and the undercount varies with age. Furthermore, the extent of the undercount varies with the census year (Schenker (1993)). This problem is again larger for blacks

than for whites.

In the death certificates, there is also error in the age at death, but this error seems to be mostly confined to blacks over the age of 65, who tend to understate their age. There is no evidence of bias in ages among whites even for those above 85 (Hill, Preston, and Rosenwaike (2000)). The overall effect of age misreporting is to downward-bias mortality for older cohorts (Preston, Elo, and Stewart (1999)).

In the absence of additional data, there is no obvious way to correct mortality rates for these problems. Overall age misreporting appears to be a very important issue mostly among blacks. These data issues suggest that our results for blacks must be taken with caution.

1.2.2 Cause of death

Another issue is whether causes of death are correctly specified in the death certificate and whether there have been significant changes from 1970 to 2000 in the accuracy with which causes of death are reported. There were two changes in the International Classification of Diseases (ICD) during our period, one in 1978 (from ICD8 to ICD9) and another in 1998 (to ICD10). These changes have affected trends in mortality rates by cause, but previous research has suggested the effects of these classification changes are small for broad causes of death such as cancer and CVD (Jemal, Ward, Anderson, and Thun (2003), Klebba (1980) and Anderson, Minio, Hoyert, and Rosenberg (2001)). Furthermore, studies that have compared the causes of death reported in the death certificate with the cause of death from an autopsy, have found that the quality of death certificate reporting has not changed much since the 1960s, except perhaps for the very old (Hoel, Ron, Carter, and Mabuchi (1993)). Overall previous research suggests that changes in the observed causes of death have not significantly changed over time for broad causes of death.

2 Competing Risks

2.1 Peterson Bounds: All demographic groups

Peterson bounds for cancer survival rates are shown in Figure 3. This figure shows the estimated bounds for all demographic groups, not only white males. It is clear from these graphs that it is not possible to draw any conclusion about the trends in cancer mortality using these bounds.

2.2 Additional Evidence of dependence between CVD and cancer

The American Heart Association lists smoking, drinking alcohol in large amounts, and obesity as factors that increase the likelihood of coronary heart disease, stroke, high blood pressure and hypertension. Moderate alcohol consumption and exercise on the other hand reduce blood pressure and coronary heart disease. The National Cancer Institute and the American Cancer Society also document that the same factors affect the risk of certain cancers. Smoking increases cancers of the respiratory system, as well as other cancers. Obesity increases the risk of cancer of the uterus, breast and prostate, among others. Excessive alcohol use increases the risk of cancer of the mouth, pharynx, larynx, esophagus, liver, and breast. Exercise is thought to reduce the risk of colon and breast cancers, and moderate alcohol consumption may lower the risk of leukemia, skin, breast and prostate cancers. This evidence suggests that at the individual level, cancer and CVD are not independent risks.

There is substantial evidence of genetic differences across individuals with respect to their susceptibility to both CVD (Nabel (2003)) and cancer (e.g. Lynch and de la Chapelle (2003), Wooster and Weber (2003)).[1] This will cause the latent duration until death from CVD and cancer to be correlated. Furthermore there are large differences in the population in terms of exposure to environmental factors and behaviors that increase particular death risks. For example in 2000, high school dropouts were more than twice as likely to smoke than college–educated individuals; women below poverty level were twice as likely as women in the highest income levels to be obese; married individuals were less likely to exercise than those who have never married; and Hispanics were less likely than non-Hispanics to drink (Schoenborn, Adams, Barnes, Vickerie, and Schiller (2004)).

Overall this evidence would suggest that the same factors that increase the likelihood of CVD also increase the risk of cancer. Evidence from country level data is also consistent with positive dependence across these two risks. For example, Vaupel and Yashin (1999) report that cancer mortality rates are lowest in countries with the highest CVD rate, suggesting that progress against CVD increases cancer mortality rates.[2]

[1] See the web pages of the American Heart Association and the National Cancer Institute for additional cites.

[2] Also it is interesting that, as in United States, mortality rates in Western Europe from cancer increased until the late 1980s and started declining thereafter; while mortality rates from CVD fell since 1970. For trends in cancer in the Europe see Levi, Negri, and La Vecchia (2000) and Levi, Negri, and La Vecchia (2002), and for trends in CVD see Levi, Negri, and La Vecchia (2002).

2.3 Accelerated failure time model and sample selection models

Honoré and Lleras-Muney (2006) consider the case where a binary explanatory variable, X, has a multiplicative effect on both of the latent distributions,

$$(T^*, I) = \begin{cases} (\min\{S_1, S_2\}, 1\{S_1 < S_2\}) & \text{for } X = 0, \\ (\min\{\alpha S_1, \beta S_2\}, 1\{\alpha S_1 < \beta S_2\}) & \text{for } X = 1, \end{cases} \tag{1}$$

where (S_1, S_2) is independent of X, and the multiplicative effects, α and β, are the main objects of interest. This model is an example of an accelerated failure time model, which is commonly used to describe mortality. It was originally introduced by Cox (1972), who gave it a physical interpretation in the context of mortality. From the equivalence between proportional hazard models with Weibull baseline hazards and Weibull accelerated failure time models, it follows that a model where the marginals obey a Weibull proportional hazard assumption is consistent with our functional form assumption. It is also a special case of the kind of general sample selection models that have been considered in the econometric literature. Specifically, if the durations are not grouped, then one can write the model in (1) as a switching regression model. See Amemiya (1985). Specifically, let $\varepsilon_k = \log(S_k)$ and consider $\log(T_1)$

$$\log(T_1) = X \cdot \log(\alpha) + \varepsilon_1$$

where $\log(T_1)$ is observed only if

$$X \cdot (\log(\beta) - \log(\alpha)) + (\varepsilon_2 - \varepsilon_1) < 0$$

The standard sufficient conditions for identification of such models require that X has "full rank" conditional on the probability that the selection criterion is satisfied (i.e. conditional on the so–called propensity score). See for example Ahn and Powell (1993). This sufficient condition is not satisfied here. Moreover, it is clear that a model with a finite number of points of support for the explanatory variable and a discrete outcome variable will not be point–identified (by the same intuition why a semiparametric discrete choice model is not identified if the explanatory variables take only a finite number of values).

2.4 Main Set-up and Alternative linear programming problem.

The identified region for (α, β) is the set of (a, b) such that there exists $p(s_1, s_2)$ satisfying

$$\sum_{\substack{t_k<s_1<t_{k+1} \\ s_2>s_1}} p(s_1, s_2) = P(T = t_k, I = 1|\, X = 0), \tag{2}$$

$$\sum_{\substack{t_k<s_2<t_{k+1} \\ s_1>s_2}} p(s_1, s_2) = P(T = t_k, I = 0|\, X = 0\,), \tag{3}$$

$$\sum_{\substack{t_k<as_1<t_{k+1} \\ bs_2>as_1}} p(s_1, s_2) = P(T = t_k, I = 1|\, X = 1\,), \tag{4}$$

$$\sum_{\substack{t_k<bs_2<t_{k+1} \\ as_1>bs_2}} p(s_1, s_2) = P(T = t_k, I = 0|\, X = 1\,), \tag{5}$$

$$\sum_{s_1,s_2} p(s_1, s_2) = 1, \qquad p(s_1, s_2) \geq 0 \tag{6}$$

(where the first four equations hold for all $k = 1, .., M$). These equations have exactly the same structure as the constraints of a linear programming problem. Specifically, for given a and b consider the linear programming problem

$$f(a, b) = \max_{\{v_i\},\{p(\cdot,\cdot)\}} \sum -v_i \tag{7}$$

subject to

$$v_k + \sum_{\substack{t_k<s_1<t_{k+1} \\ s_2>s_1}} p(s_1, s_2) = P(T = t_k, I = 1|\, X = 0) \qquad k = 1, \ldots M, \tag{8}$$

$$v_{M+k} + \sum_{\substack{t_k<s_2<t_{k+1} \\ s_1>s_2}} p(s_1, s_2) = P(T = t_k, I = 0|\, X = 0\,) \qquad k = 1, \ldots M, \tag{9}$$

$$v_{2M+k} + \sum_{\substack{t_k<as_1<t_{k+1} \\ bs_2>as_1}} p(s_1, s_2) = P(T = t_k, I = 1|\, X = 1\,) \qquad k = 1, \ldots M, \tag{10}$$

$$v_{3M+k} + \sum_{\substack{t_k<bs_2<t_{k+1} \\ as_1>bs_2}} p(s_1, s_2) = P(T = t_k, I = 0|\, X = 1\,) \qquad k = 1, \ldots M, \tag{11}$$

$$v_{4M+1} + \sum_{s_1,s_2} p(s_1, s_2) = 1, \qquad p(s_1, s_2) \geq 0 \qquad \text{for all } (s_1, s_2), \tag{12}$$

$$v_i \geq 0 \qquad k = 1, \ldots 4M + 1 \tag{13}$$

Note that the setup in (7) forces one to underestimate all the probabilities. While this does not affect the consistency of the resulting estimator of α and β, it may be intuitively unappealing. It might therefore be more attractive to consider the linear programming problem

$$f(a,b) = \max_{\{v_i\},\{u_i\},\{p(\cdot,\cdot)\}} \sum -(v_i + u_i) \tag{14}$$

subject to

$$\begin{aligned}
v_k - u_k + \sum_{\substack{t_k<s_1<t_{k+1} \\ s_2>s_1}} p(s_1,s_2) &= P(T=t_k, I=1 | X=0) \qquad k=1,\dots M, \\
v_{M+k} - u_{M+k} + \sum_{\substack{t_k<s_2<t_{k+1} \\ s_1>s_2}} p(s_1,s_2) &= P(T=t_k, I=0 | X=0\ ) \qquad k=1,\dots M, \\
v_{2M+k} - u_{2M+k} + \sum_{\substack{t_k<as_1<t_{k+1} \\ bs_2>as_1}} p(s_1,s_2) &= P(T=t_k, I=1 | X=1\ ) \qquad k=1,\dots M, \\
v_{3M+k} - u_{3M+k} + \sum_{\substack{t_k<bs_2<t_{k+1} \\ as_1>bs_2}} p(s_1,s_2) &= P(T=t_k, I=0 | X=1\ ) \qquad k=1,\dots M, \\
v_{4M+1} - u_{4M+1} + \sum_{s_1,s_2} p(s_1,s_2) &= 1, \\
p(s_1,s_2) &\geq 0 \qquad \text{for all } (s_1,s_2), \\
u_i, v_i &\geq 0 \qquad k=1,\dots 4M+1.
\end{aligned}$$

The disadvantage of this approach is that it increases the dimensionality of the linear programming problem. In our application, we therefore focus on the first formulation.

3 Trends in Cardiovascular and Cancer mortality (1970-2000)

3.1 Details about the Calculations

3.1.1 Grid search

The function value that defines the identified region was calculated over three grids.

The first grid was defined by the rectangle $\{0.90, 0.95, 1.00, ..., 1.40\} \times \{0.90, 0.95, 1.00, ..., 1.40\}$.

The second grid was defined by first calculating the set of maximizers over the original grid. Let $\theta_1^{\min}$ and $\theta_1^{\max}$ denote the minimum and maximum value of the first coordinate in that set and let $\theta_2^{\min}$ and $\theta_2^{\max}$ denote the minimum and maximum value of the second coordinate in the set. The second grid is then given by $\left\{\theta_1^{\min} - 0.05, \theta_1^{\min} - 0.04, \theta_1^{\min} - 0.03, ..., \theta_1^{\max} + 0.08\right\} \times \left\{\theta_2^{\min} - 0.05, \theta_2^{\min} - 0.04, \theta_2^{\min} - 0.03, ..., \theta_2^{\max} + 0.08\right\}$.

The third grid was defined in terms of the maximizers over the first two grids. Let $\theta_1^{\min}$ and $\theta_1^{\max}$ denote the minimum and maximum value of the first coordinate in that set and let $\theta_2^{\min}$ and $\theta_2^{\max}$ denote the minimum and maximum value of the second coordinate in the set. The third grid

is then given by $\left\{\theta_1^{\min}-0.01, \theta_1^{\min}-0.009, \theta_1^{\min}-0.008, ..., \theta_1^{\max}+0.015\right\} \times$
$\left\{\theta_2^{\min}-0.01, \theta_2^{\min}-0.009, \theta_2^{\min}-0.008, ..., \theta_2^{\max}+0.015\right\}$.

The estimated identified region is the set of maximizers of the union of the three grids. The numbers reported in the tables are the minimum and maximum values of each coordinate.

3.1.2 Choice of ε_n

The theory presented earlier requires one to define the interval estimates as the set of parameter values for which the function value is within some ε_n of its maximum, where ε_n is a sequence of numbers that converges to 0 more slowly than $\sup_{a,b}\left|\widehat{f}(a,b)-f(a,b)\right|$. Since measurement error is likely to be more important than estimation uncertainty, we ignore this issue in the results presented in the paper. Consider, for example the improvement for white males between 1970 and 2000. This is based on approximate 40,000,000 observations in each of the two years. A small simulation study that draws observations using the sample probabilities, suggests that the average sup–difference is approximately 0.00057 with a standard deviation of approximately 0.00016 (the maximum value over 1000 replications was 0.001). If we choose ε_n to be the mean plus five standard deviations (0.00137), then the interval estimates for a and b would change to $(1.34, 1.40)$ and $(1.11, 1.25)$, respectively. While these intervals are somewhat larger than the intervals presented in Table 3, the substantive conclusions based on them are not different.

3.1.3 Estimation under independence

To estimate the parameters under independence, we first estimate the marginal distribution of T_1 and T_2 using a Kaplan–Meier estimator. We then estimate bounds for a by

$$f(a) = \max_{\{v_i\},\{p(\cdot,\cdot)\}} \sum -v_i$$

subject to

$$\begin{aligned}
v_k + \sum_{\substack{t_k<s_1<t_k+1 \\ s_2>s_1}} p(s_1) &= \widehat{P}(T=t_k, I=1|\, X=0) \qquad k=1,\ldots M, \\
v_{M+k} + \sum_{\substack{t_k<as_1<t_k+1 \\ bs_2>as_1}} p(s_1) &= \widehat{P}(T=t_k, I=1|\, X=1\,) \qquad k=1,\ldots M, \\
v_{2M+1} + \sum_{s_1,s_2} p(s_1) &= 1, \\
p(s_1) &\geq 0 \qquad \text{for all } s_1, \\
v_i &\geq 0 \qquad k=1,\ldots 2M+1
\end{aligned}$$

where the points of support are determined in a way that is similar to the way we did it without independence. Bounds for b are estimated analogously.

3.2 Main results: Additional evidence

Our findings provide support for the claim that there has been progress in cancer, measured in terms of the increases in the underlying cause-specific duration. We looked for any evidence that there were indeed innovations in terms of cancer treatment during the period we study, starting in the 1970s for women and mostly in the 1990s for men. We focus on improvements for the major cancer sites (excluding lung[3]), namely breast, prostate, colorectal and ovarian cancer.

Survival from colorectal cancer, which disproportionately affects men, has improved because of a combination of earlier detection and improved treatment at earlier stages. Standard treatment for colorectal cancer changed in 1990, following a National Institutes of Health Conference recommendation, to include a combination of 5FU and leucovorin, two previously existing drugs (NIH Consensus Conference (1990)). Although treatment for prostate cancer remains controversial, clinical trials in the 1990s showed promising effects of hormonal treatment (Howe, Wingo, Thun, Ries, Rosenberg, Feigal, and Edwards (2001)).

Improvements to treat women's cancers started earlier. Mammographies started being routinely offered in the 1970s and studies in the 1970s and 1980s showed that early detection substantially improved mortality, especially for women over 50.[4] Breast cancer treatment also changed in the 1980s with the dissemination of adjuvant chemotherapy, including multi-agent chemotherapy and tamoxifen. Additional changes in treatment were implemented in the early 1990s for postmenopausal women (Mariotto, Feuer, Harlan, Wun, Johnson, and Abrams (2002)). Treatment for ovarian cancer was modified in 1986 (NIH Consensus Conference (1995)) to include surgery and chemotherapy with a platinum compound (cisplatin or carboplatin) after publication of results from randomized trials which showed their effectiveness (Omura, Blessing, Ehrlich, Miller, Yordan, Creasman, and Homesley (1986)).

Overall this evidence from other sources is consistent with our findings on the timing of progress against cancer by gender.

[3]The fight against lung cancer has mostly focused on reducing tobacco consumption. This effort began with the Surgeon General Report in 1964 that first publicly announced that smoking increased the risk of lung cancer, and continues today. These efforts are reflected in the trends in lung cancer many years later. To our knowledge there is no evidence of other forms of progress in lung cancer.

[4]A review of the evidence by the U.S. Preventive Services Task Force is available at http://www.ahrq.gov/clinic/3rduspstf/breastcancer/brcanrr.htm#ref4

3.3 Policy applications: Counterfactuals

The results in Table 4a report counterfactuals for every decade, rather than just 1970-2000.

The bounds on the changes implied by the results in Table 4 of Honoré and Lleras-Muney (2006) and Table 4a are potentially too wide as they are based on a comparison of bounds for two parameters. Alternatively we can estimate bounds on the change directly. These alternative bound estimates are in Table 5. The bounds that are estimated directly are somewhat tighter, although the general pattern and the qualitative conclusions (reported in the paper) do not change.

We also use our model to calculate changes in life expectancy conditional on survival to age 45 (and with censoring at 80). The results, presented in Table 6, are very similar to those in Table 4, except that we can express changes in the survival distribution in terms of additional years of life. These results can be used in cost benefit analysis if we have estimates of the dollar value of an additional year of life. For example, for white males the actual increase from 1970 to 2000 is approximately 3.3 years (the fitted value are between 2.96 and 3.13). In the absence of progresses in cancer, progress in CVD disease would account for 0.94 to 2.73 years. One estimate of the progress in cancer is therefore the remaining 0.3 to 2.1 years of life. As of 2000, given the progress that occurred in CVD, eliminating cancer would further increase life expectancy by an additional 0.3 to 1.37 years. For white females, the potential gains from eliminating cancer in 2000 are smaller, 0 to 1.2 years. This is somewhat unintuitive since cancer is a relatively larger risk for females than for males. However this is partially due to the fact that the parameter of interest is lifetime censored at 80. Since whites females live the longest (see Table 1) the censoring gives rise to lower progress estimates for women. For this reason we prefer the estimates using the change in the probability of survival past age 75.

3.4 Specification Checks

3.4.1 Functional form

Tables 7a and 7b present the estimated coefficients from the model that allows the trends for the later age groups to differ as described in section 6 of Honoré and Lleras-Muney (2006).

We use the coefficients from these tables to construct Figure 4 of Honoré and Lleras-Muney (2006) which reports the results for white males. Figures 4 and 5 use the estimates to construct similar figures for our four demographic groups and for both causes of death. These figures show that allowing for different trends by age does not substantively alter the main conclusions of the paper.

3.4.2 Excluding lung cancer

In Table 8, we present the bounds for cancer excluding lung cancer with and without assuming independence. We find much larger improvements when we exclude lung cancer for all groups (compare the results in this table to those in Table 3 of the paper). The trends are about twice as large as those that include lung cancer, about 19% and 46% for white men and women respectively, and 9% and 45% for black men and women. As the table shows these improvements are also much larger than those we obtain when we assume independence.

3.4.3 No functional form assumption on CVD

We estimate bounds for cancer that impose a multiplicative effect on cancer only. These results are presented in Table 9. In all cases, relaxing the parametric assumption for CVD results in bounds that are very large, typically ranging from about 0.5 to about 2.3. It is therefore difficult to draw any conclusions from these results.

3.4.4 Starting at age 40

If progress in detection and treatment prevents some deaths prior to age 45 in 2000 but not in 1970, then our choice to restrict the sample to those ages 45 and above could bias our results. There are two reasons why we present results only conditional on survival until age 45. The first is that cancers that affect children and young adults are considered to be very different from those that affect mature adults. The trends for cancer would therefore be expected to be quite different for these age groups. The second and main reason for the age restriction is that the death rates below age 45 were low for both causes of death in all years (see Figures 6 and 7). However the graphs suggest that mortality rates started rising between ages 40 and 45. Thus we re-estimated the results in Table 3 starting with ages 40 (instead of 45). The results are presented in Table 10. Overall, we find substantial progress in cancer from 1970 to 2000, and the timing of this progress is as described in the main result section.

References

Ahn, H., and J. L. Powell (1993): "Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism," *Journal of Econometrics*, 58(1-2), 3–29.

Amemiya, T. (1985): *Advanced Econometrics*. Harvard University Press.

Anderson, R. N., A. M. Minio, D. L. Hoyert, and H. M. Rosenberg (2001): "Comparability of Causes of Death Between ICD-9 and ICD-10: Preliminary Estimates," *National Vital Statistics Reports*, 49(2).

Cox, D. R. (1972): "Regression models and life-tables," *J. Royal Statistical Soc. Series B-statistical Methodology*, 34.

Hill, M. E., S. H. Preston, and I. Rosenwaike (2000): "Age Reporting Among White Americans Aged 85+: Results Of A Record Linkage Study," *Demography*, 37(2).

Hoel, D. G., E. Ron, R. Carter, and K. Mabuchi (1993): "Influence of Death Certificate Errors on Cancer Mortality Trends," *Journal of the National Cancer Institute*, 85(13), 1063–8.

Honoré, B. E., and A. Lleras-Muney (2006): "Bounds in Competing Risks and the War on Cancer," submitted to *Econometrica*.

Howe, H., P. Wingo, M. J. Thun, L. A. Ries, H. M. Rosenberg, E. G. Feigal, and B. K. Edwards (2001): "Annual Report to the Nation on the Status of Cancer (1973 through 1998), Featuring Cancers with Recent Increasing Trends," *Journal of the National Cancer Institute*, 93(11).

Jemal, A., E. Ward, R. N. Anderson, and M. J. Thun (2003): "Influence of Rules From the Tenth Revision of the International Classification of Diseases on U.S. Cancer Mortality Trends," *Journal of the National Cancer Institute*, 95, 1727–1728.

Klebba, A. J. (1980): "Estimates of Selected Comparability Ratios Based on Dual Coding of the 1976 Death Certificates by the Eighth and Ninth Revision of the International Classification of Diseases," *Monthly Vital Statistics Report*, 28(11).

Levi, F., L., F. E. Negri, and C. La Vecchia (2000): "The Decline in Cancer Mortality in the European Union," *European Journal of Cancer*, 36(15), 1965–1968.

——— (2002): "Trends in Mortality from Cardiovascular and Cerebrovascular Diseases in Europe and Other Areas of the World," *Heart*, 88, 119–124.

Lynch, H. T., and A. de la Chapelle (2003): "Hereditary colorectal cancer," *New England Journal of Medicine*, 348, 919–932.

Mariotto, A., E. J. Feuer, L. C. Harlan, L.-M. Wun, K. A. Johnson, and J. Abrams (2002): "Trends in Use of Adjuvant Multi–Agent Chemotherapy and Tamoxifen for Breast Cancer in the United States 1975-1999," *Journal of the National Cancer Institute*, 94(21).

Nabel, E. G. (2003): "Genomic Medicine: Cardiovascular Disease," *New England Journal of Medicine*, 349, 60–72.

NIH Consensus Conference (1990): "Adjuvant Therapy for Patients with Colon and Rectal Cancer," *Journal of the American Medical Association*, 264, 1444–1450.

——— (1995): "Ovarian Cancer, Screening, Treatment and Followup," *Journal of the American Medical Association*, 273, 491–497.

Omura, G., J. A. Blessing, C. E. Ehrlich, A. Miller, E. Yordan, W. T. Creasman, and H. D. Homesley (1986): "A Randomized Trial of Cyclophosphamide abdand Doxorubicin with or Without Cisplatin in Advanced Ovarian Carcinoma: A Gynecologic Oncology Group Study," *Cancer*, 57, 1725–1730.

Preston, S. H., I. T. Elo, and Q. Stewart (1999): "Effects of Age Misreporting on Mortality Estimates at Older Ages," *Population Studies*, 53(2), 165–177.

Schenker, N. (1993): "Undercount in the 1990 Census: Special Section," *Journal of the American Statistical Association*, 88(423).

Schoenborn, C., P. F. Adams, P. M. Barnes, J. L. Vickerie, and J. S. Schiller (2004): "Health Behaviors of Adults: United States, 1999–2001," *National Center for Health Statistics. Vital Health Stat*, 10(219).

Vaupel, J. W., and A. I. Yashin (1999): "Cancer Rates over Age, Time and Place: Insights from Stochastic Models of Heterogenous Populations," *Max Plank Institute for Demographic Reseach Working Paper 1999-006*.

Wooster, R., and B. L. Weber (2003): "Genomic Medicine: Breast and Ovarian Cancer," *New England Journal of Medicine*, 348, 2339–2347.

Figure 1: Hazard Rates for the Cancer

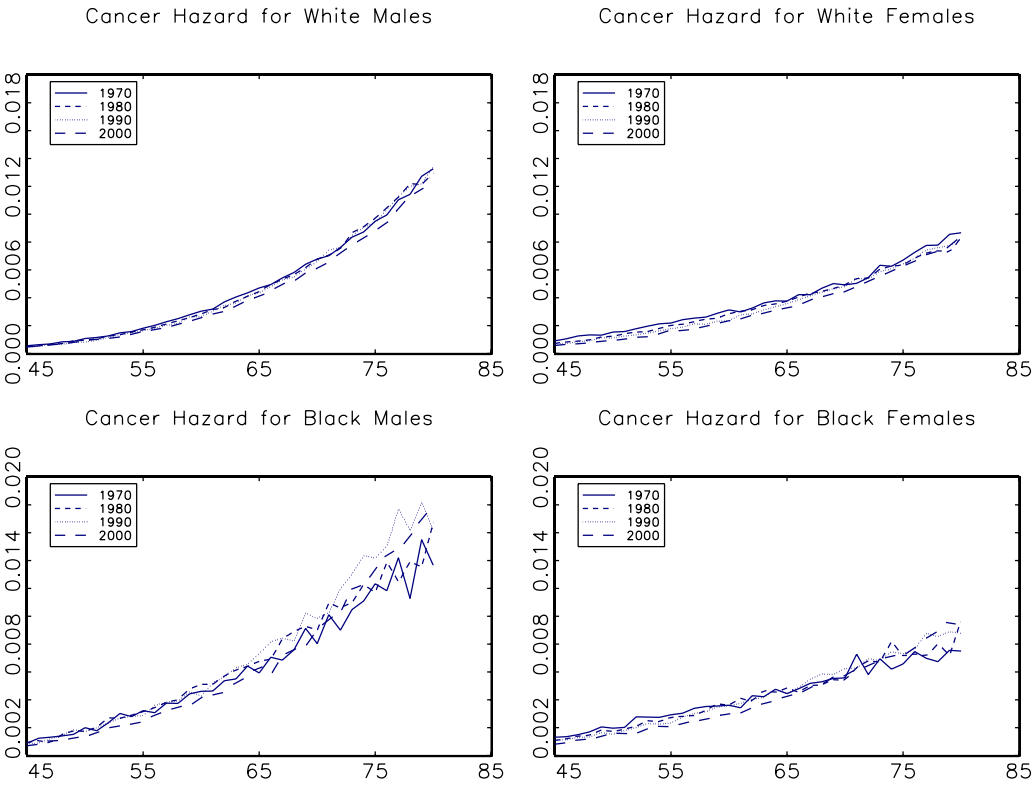

Figure 2: Hazard Rates for the CVD

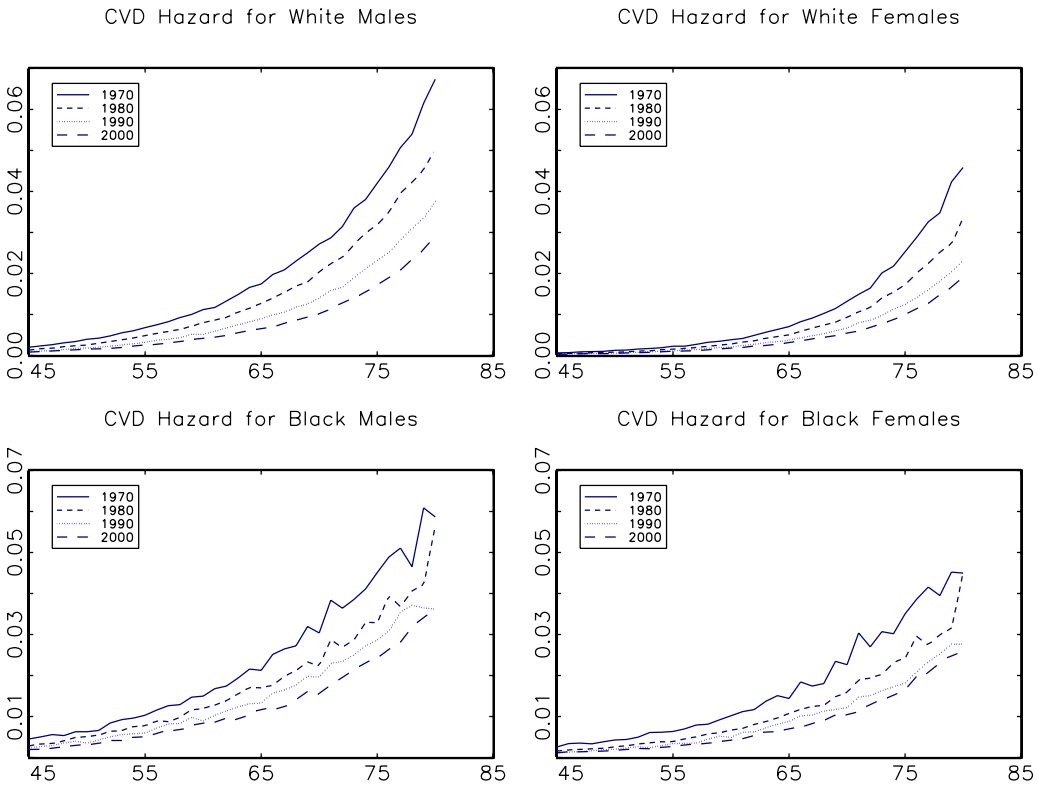

Figure 3: Peterson Bounds on the Cancer Survivor Functions in 1970 and 2000

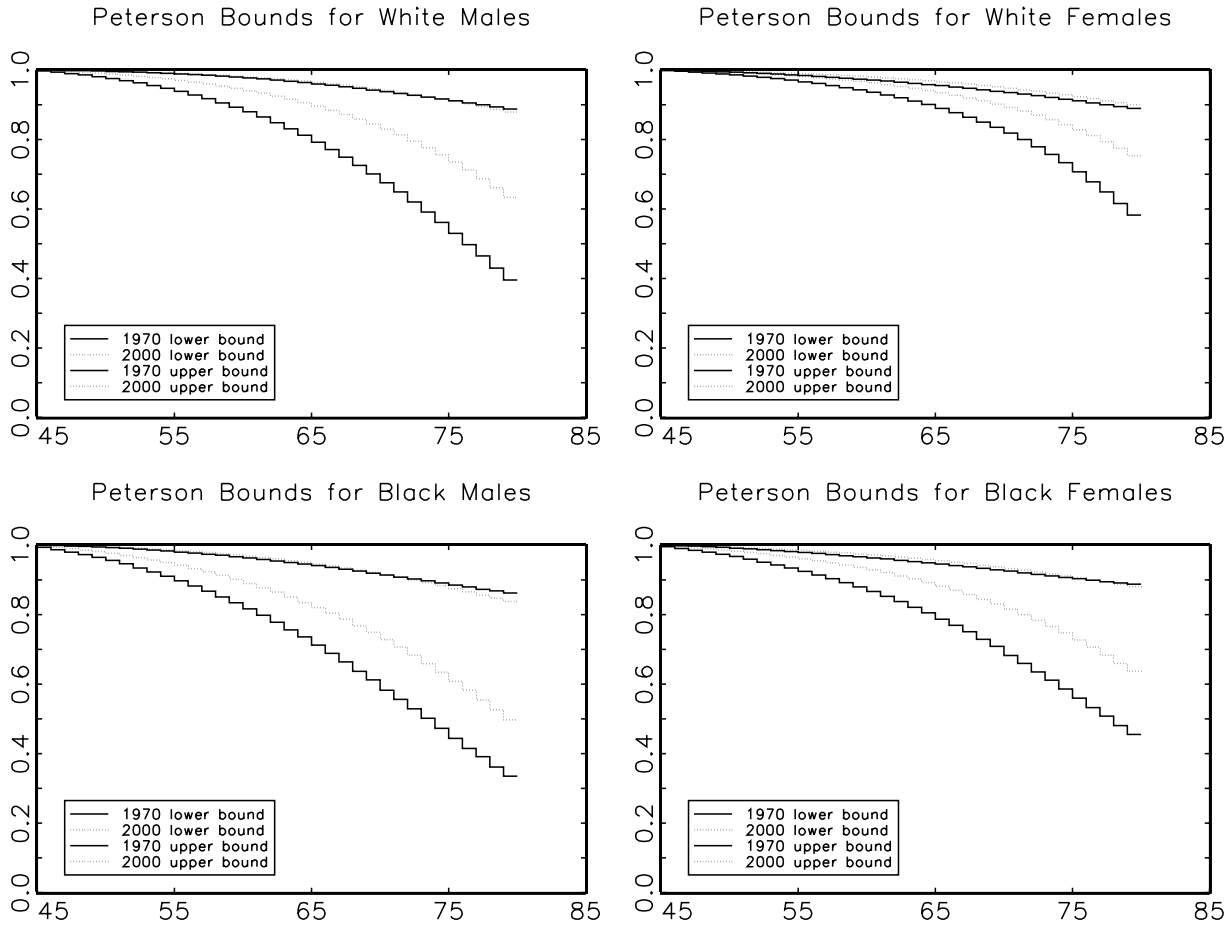

Figure 4: Allowing for Kinks: CVD

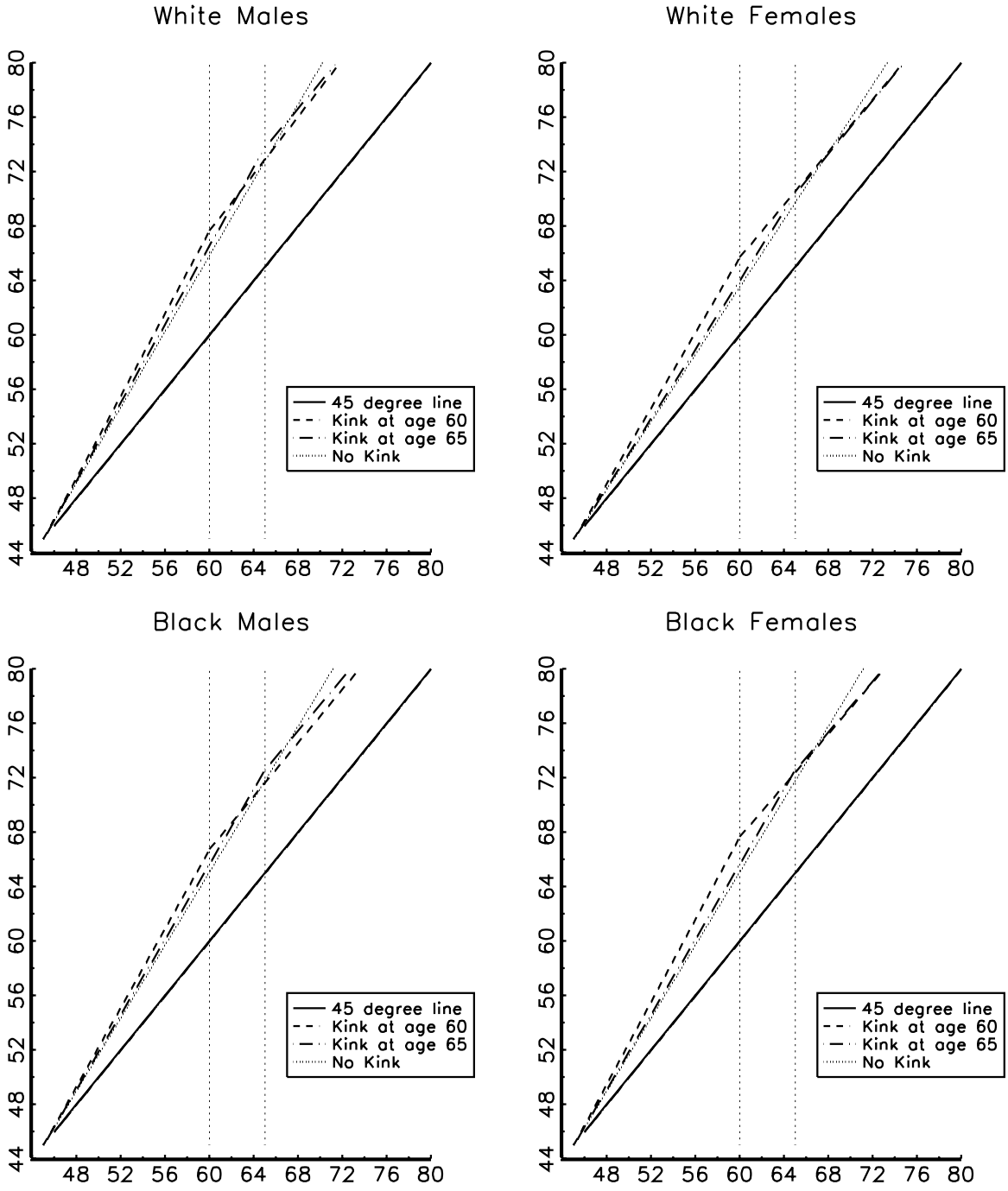

Figure 5: Allowing for Kinks: Cancer

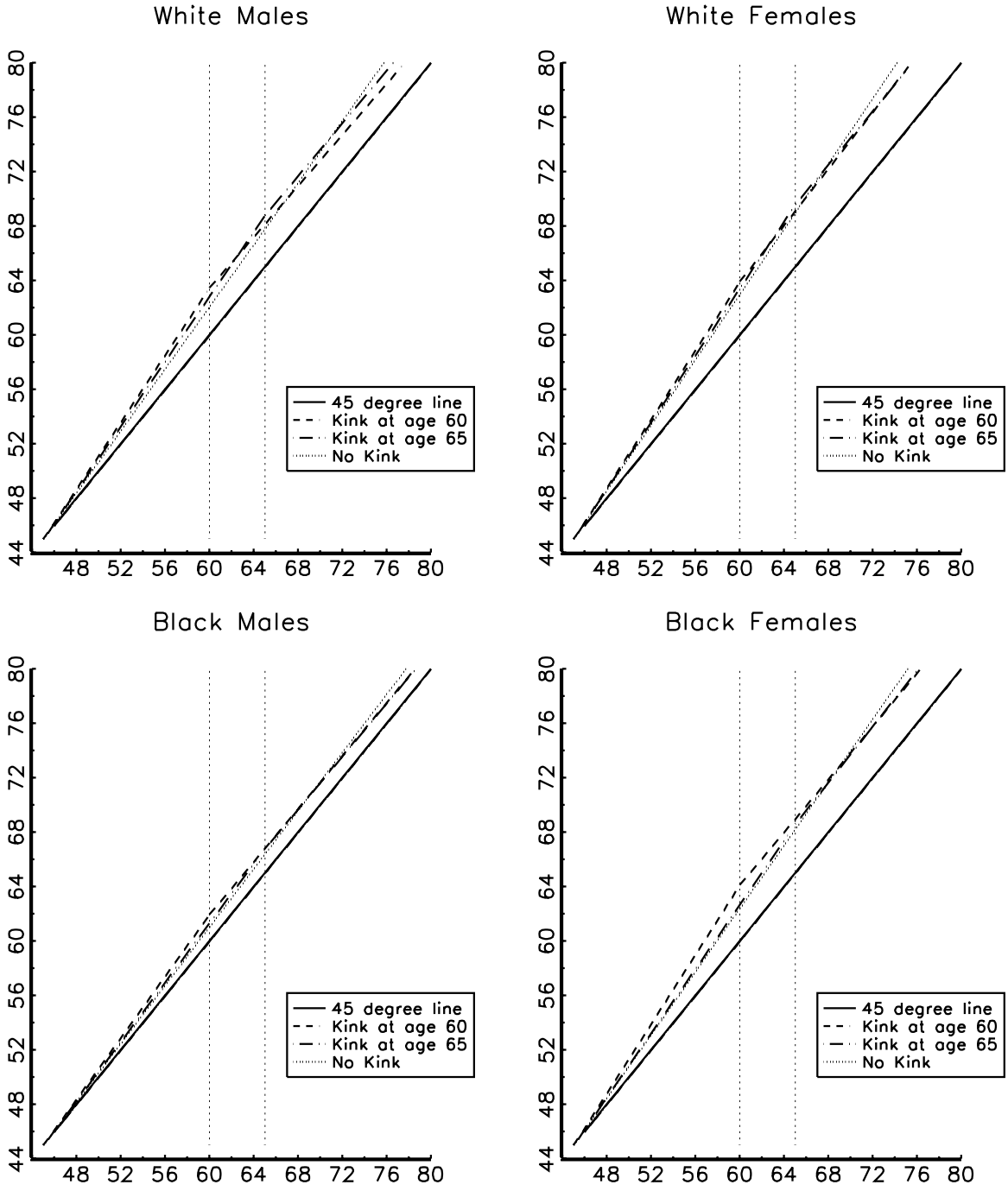

Figure 6: CVD Death Rates by Age in 1970

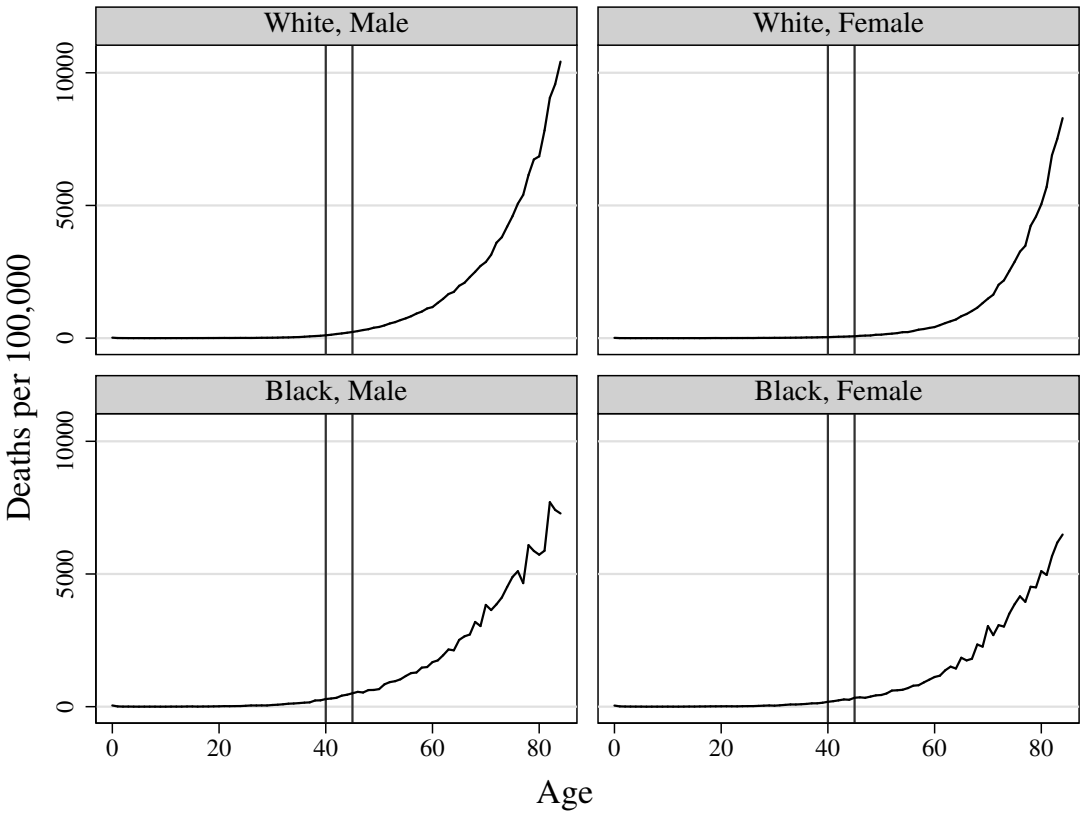

Figure 7: Cancer Death Rates by Age in 1970

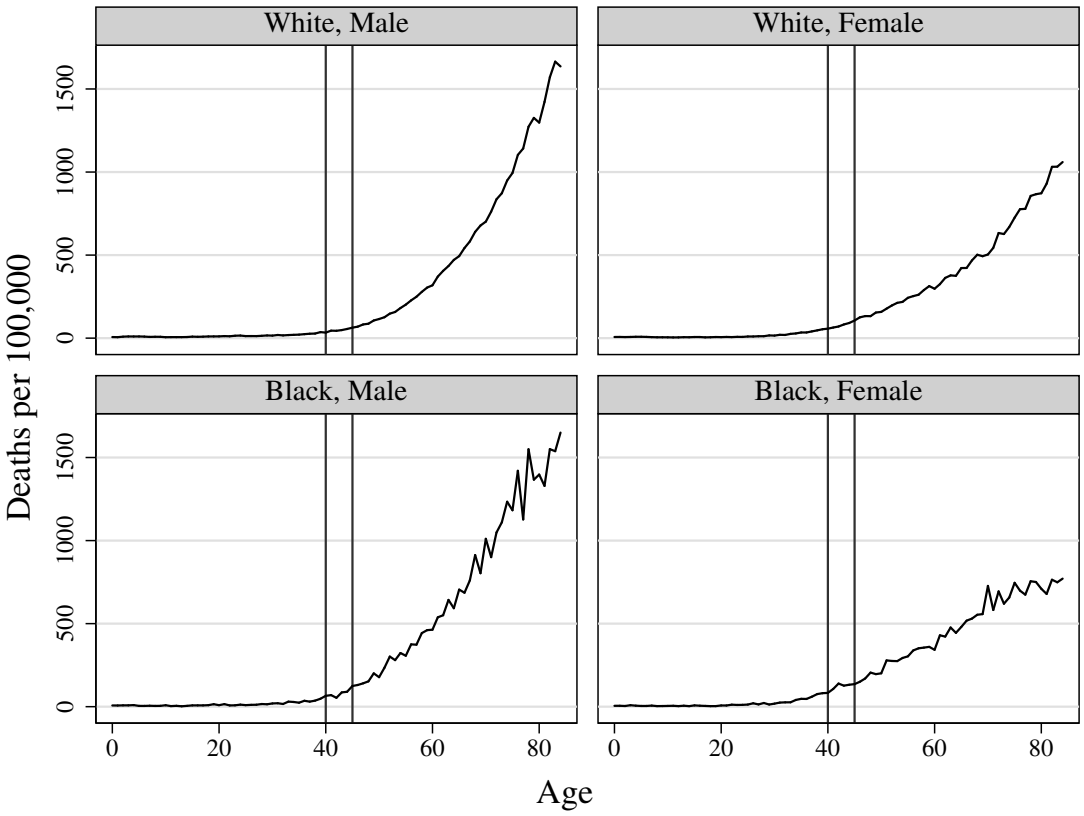

TABLE 4a: Counterfactual Probability of Surviving Age 75

	1970–1980	1970–1990	1970–2000
White Males			
Change in actual prob.	$0.561 - 0.636$	$0.561 - 0.707$	$0.561 - 0.756$
No Progress (fitted in 70)	$(0.567, 0.567)$	$(0.572, 0.573)$	$(0.574, 0.574)$
Progress in CVD	$(0.634, 0.643)$	$(0.699, 0.719)$	$(0.661, 0.738)$
Progress in Cancer	$(0.567, 0.572)$	$(0.572, 0.579)$	$(0.574, 0.595)$
Progress in Both (fitted in end year)	$(0.638, 0.643)$	$(0.707, 0.719)$	$(0.756, 0.769)$
Elim. CVD — Cancer as in end year	$(0.638, 0.913)$	$(0.707, 0.906)$	$(0.756, 0.919)$
Elim. Cancer — CVD as in end year	$(0.661, 0.732)$	$(0.747, 0.812)$	$(0.785, 0.848)$
Elim. CVD — Cancer as in 70	$(0.634, 0.909)$	$(0.699, 0.906)$	$(0.661, 0.888)$
Elim. Cancer — CVD as in 70	$(0.567, 0.656)$	$(0.572, 0.662)$	$(0.574, 0.663)$
White Females			
Change in actual prob.	$0.733 - 0.784$	$0.733 - 0.820$	$0.733 - 0.843$
No Progress (fitted in 70)	$(0.736, 0.736)$	$(0.738, 0.738)$	$(0.739, 0.740)$
Progress in CVD	$(0.736, 0.776)$	$(0.738, 0.802)$	$(0.739, 0.820)$
Progress in Cancer	$(0.736, 0.751)$	$(0.738, 0.760)$	$(0.739, 0.765)$
Progress in Both (fitted in end year)	$(0.786, 0.787)$	$(0.824, 0.826)$	$(0.843, 0.850)$
Elim. CVD — Cancer as in end year	$(0.786, 0.917)$	$(0.824, 0.923)$	$(0.843, 0.930)$
Elim. Cancer — CVD as in end year	$(0.786, 0.865)$	$(0.824, 0.900)$	$(0.843, 0.916)$
Elim. CVD — Cancer as in 70	$(0.736, 0.906)$	$(0.738, 0.900)$	$(0.739, 0.900)$
Elim. Cancer — CVD as in 70	$(0.736, 0.824)$	$(0.738, 0.827)$	$(0.739, 0.827)$
Black Males			
Change in actual prob.	$0.473 - 0.540$	$0.473 - 0.577$	$0.473 - 0.634$
No Progress (fitted in 70)	$(0.474, 0.474)$	$(0.481, 0.482)$	$(0.486, 0.486)$
Progress in CVD	$(0.513, 0.542)$	$(0.536, 0.579)$	$(0.582, 0.629)$
Progress in Cancer	$(0.474, 0.481)$	$(0.481, 0.490)$	$(0.486, 0.504)$
Progress in Both (fitted in end year)	$(0.542, 0.542)$	$(0.587, 0.588)$	$(0.635, 0.647)$
Elim. CVD — Cancer as in end year	$(0.542, 0.872)$	$(0.587, 0.871)$	$(0.635, 0.880)$
Elim. Cancer — CVD as in end year	$(0.558, 0.660)$	$(0.621, 0.708)$	$(0.678, 0.768)$
Elim. CVD — Cancer as in 70	$(0.513, 0.872)$	$(0.536, 0.862)$	$(0.582, 0.862)$
Elim. Cancer — CVD as in 70	$(0.474, 0.589)$	$(0.481, 0.597)$	$(0.486, 0.599)$
Black Females			
Change in actual prob.	$0.586 - 0.673$	$0.586 - 0.713$	$0.586 - 0.748$
No Progress (fitted in 70)	$(0.594, 0.594)$	$(0.598, 0.598)$	$(0.603, 0.603)$
Progress in CVD	$(0.616, 0.673)$	$(0.598, 0.696)$	$(0.668, 0.740)$
Progress in Cancer	$(0.594, 0.604)$	$(0.598, 0.615)$	$(0.603, 0.622)$
Progress in Both (fitted in end year)	$(0.679, 0.683)$	$(0.717, 0.725)$	$(0.748, 0.764)$
Elim. CVD — Cancer as in end year	$(0.679, 0.907)$	$(0.717, 0.909)$	$(0.748, 0.916)$
Elim. Cancer — CVD as in end year	$(0.690, 0.774)$	$(0.726, 0.807)$	$(0.768, 0.846)$
Elim. CVD — Cancer as in 70	$(0.616, 0.897)$	$(0.598, 0.881)$	$(0.668, 0.891)$
Elim. Cancer — CVD as in 70	$(0.594, 0.691)$	$(0.598, 0.694)$	$(0.603, 0.698)$

TABLE 5: Bound on the Change in Counterfactual Probability of Surviving Age 75

	1970–1980	1970–1990	1970–2000
White Males			
Increase in prob. (w. impr in CVD)	$(0.000, 0.027)$	$(0.000, 0.033)$	$(0.027, 0.108)$
Increase in prob. (no impr in CVD)	$(0.000, 0.005)$	$(0.000, 0.011)$	$(0.000, 0.022)$
White Females			
Increase in prob. (w. impr in CVD)	$(0.010, 0.051)$	$(0.021, 0.087)$	$(0.029, 0.111)$
Increase in prob. (no impr in CVD)	$(0.000, 0.015)$	$(0.000, 0.021)$	$(0.000, 0.025)$
Black Males			
Increase in prob. (w. impr in CVD)	$(0.000, 0.029)$	$(0.009, 0.051)$	$(0.016, 0.065)$
Increase in prob. (no impr in CVD)	$(0.000, 0.006)$	$(0.000, 0.008)$	$(0.000, 0.018)$
Black Females			
Increase in prob. (w. impr in CVD)	$(0.009, 0.067)$	$(0.028, 0.127)$	$(0.022, 0.097)$
Increase in prob. (no impr in CVD)	$(0.000, 0.010)$	$(0.000, 0.017)$	$(0.000, 0.019)$

TABLE 6: Counterfactual (Censored)
Life Expectancy at age 45

	1970–1980	1970–1990	1970–2000
White Males			
Change in actual $E[T\|\, T \geq 45]$.	$72.73 - 74.09$	$72.73 - 75.28$	$72.73 - 76.05$
No Progress (fitted in 70)	$(72.66, 72.71)$	$(72.61, 72.66)$	$(72.63, 72.68)$
Progress in CVD	$(73.46, 73.93)$	$(74.26, 74.95)$	$(73.62, 75.36)$
Progress in Cancer	$(72.66, 72.81)$	$(72.61, 72.78)$	$(72.63, 73.07)$
Progress in Both (fitted in end year)	$(73.89, 73.98)$	$(74.92, 74.98)$	$(75.64, 75.76)$
Elim. CVD — Cancer as in end year	$(73.89, 78.33)$	$(74.92, 78.05)$	$(75.64, 78.22)$
Elim. Cancer — CVD as in end year	$(74.15, 75.42)$	$(75.49, 76.49)$	$(76.06, 77.01)$
Elim. CVD — Cancer as in 70	$(73.46, 78.27)$	$(74.26, 78.02)$	$(73.62, 77.82)$
Elim. Cancer — CVD as in 70	$(72.66, 74.19)$	$(72.61, 74.14)$	$(72.63, 74.16)$
White Females			
Change in actual $E[T\|\, T \geq 45]$.	$75.67 - 76.51$	$75.67 - 77.09$	$75.67 - 77.46$
No Progress (fitted in 70)	$(75.65, 75.66)$	$(75.63, 75.65)$	$(75.61, 75.63)$
Progress in CVD	$(75.65, 76.27)$	$(75.63, 76.60)$	$(75.68, 76.83)$
Progress in Cancer	$(75.65, 75.91)$	$(75.63, 76.02)$	$(75.61, 76.07)$
Progress in Both (fitted in end year)	$(76.43, 76.47)$	$(76.92, 76.97)$	$(77.25, 77.31)$
Elim. CVD — Cancer as in end year	$(76.43, 78.47)$	$(76.92, 78.46)$	$(77.25, 78.54)$
Elim. Cancer — CVD as in end year	$(76.43, 77.85)$	$(76.92, 78.21)$	$(77.30, 78.44)$
Elim. CVD — Cancer as in 70	$(75.65, 78.27)$	$(75.63, 78.10)$	$(75.68, 78.08)$
Elim. Cancer — CVD as in 70	$(75.65, 77.23)$	$(75.63, 77.21)$	$(75.61, 77.16)$
Black Males			
Change in actual $E[T\|\, T \geq 45]$.	$70.69 - 71.98$	$70.69 - 72.76$	$70.69 - 73.80$
No Progress (fitted in 70)	$(70.68, 70.69)$	$(70.61, 70.67)$	$(70.62, 70.66)$
Progress in CVD	$(71.39, 71.91)$	$(71.51, 72.42)$	$(72.19, 73.26)$
Progress in Cancer	$(70.68, 70.82)$	$(70.61, 70.90)$	$(70.62, 70.93)$
Progress in Both (fitted in end year)	$(71.93, 71.97)$	$(72.48, 72.63)$	$(73.39, 73.52)$
Elim. CVD — Cancer as in end year	$(71.93, 77.73)$	$(72.48, 77.49)$	$(73.39, 77.54)$
Elim. Cancer — CVD as in end year	$(72.22, 74.00)$	$(72.93, 74.62)$	$(74.08, 75.46)$
Elim. CVD — Cancer as in 70	$(71.39, 77.67)$	$(71.51, 77.27)$	$(72.19, 77.28)$
Elim. Cancer — CVD as in 70	$(70.68, 72.73)$	$(70.61, 72.70)$	$(70.62, 72.65)$
Black Females			
Change in actual $E[T\|\, T \geq 45]$.	$72.83 - 74.45$	$72.83 - 75.15$	$72.83 - 75.75$
No Progress (fitted in 70)	$(72.71, 72.75)$	$(72.77, 72.80)$	$(72.72, 72.77)$
Progress in CVD	$(73.09, 74.04)$	$(73.01, 74.47)$	$(73.45, 74.96)$
Progress in Cancer	$(72.71, 72.98)$	$(72.77, 73.17)$	$(72.72, 73.13)$
Progress in Both (fitted in end year)	$(74.16, 74.26)$	$(74.80, 74.90)$	$(75.22, 75.35)$
Elim. CVD — Cancer as in end year	$(74.16, 77.96)$	$(74.80, 77.93)$	$(75.22, 77.89)$
Elim. Cancer — CVD as in end year	$(74.29, 75.90)$	$(74.94, 76.41)$	$(75.53, 76.82)$
Elim. CVD — Cancer as in 70	$(73.09, 77.73)$	$(73.01, 77.51)$	$(73.45, 77.51)$
Elim. Cancer — CVD as in 70	$(72.71, 74.56)$	$(72.77, 74.61)$	$(72.72, 74.51)$

TABLE 7A: Marginal Identified Regions (break at age 60)

	1970–1980	1970–1990	1970–2000
White Males			
Coefficient on CVD before age 60	$(1.167, 1.197)$	$(1.377, 1.381)$	$(1.503, 1.528)$
Coefficient on CVD after age 60	$(0.950, 1.014)$	$(0.974, 1.053)$	$(1.021, 1.100)$
Coefficient on Cancer before age 60	$(1.052, 1.111)$	$(1.170, 1.195)$	$(1.223, 1.267)$
Coefficient on Cancer after age 60	$(0.906, 0.956)$	$(0.954, 1.033)$	$(1.009, 1.088)$
Function Value	-0.001073	-0.002334	-0.002246
White Females			
Coefficient on CVD before age 60	$(1.169, 1.194)$	$(1.300, 1.308)$	$(1.377, 1.381)$
Coefficient on CVD after age 60	$(0.952, 1.021)$	$(0.963, 1.023)$	$(0.952, 1.016)$
Coefficient on Cancer before age 60	$(1.128, 1.132)$	$(1.184, 1.194)$	$(1.255, 1.260)$
Coefficient on Cancer after age 60	$(0.951, 1.011)$	$(0.989, 1.058)$	$(1.011, 1.085)$
Function Value	-0.000797	-0.001479	-0.002890
Black Males			
Coefficient on CVD before age 60	$(1.147, 1.176)$	$(1.301, 1.306)$	$(1.457, 1.457)$
Coefficient on CVD after age 60	$(1.023, 1.052)$	$(0.969, 1.028)$	$(0.940, 1.014)$
Coefficient on Cancer before age 60	$(1.004, 1.029)$	$(1.004, 1.019)$	$(1.126, 1.126)$
Coefficient on Cancer after age 60	$(0.936, 1.015)$	$(0.956, 1.015)$	$(0.944, 1.013)$
Function Value	-0.000768	-0.001481	-0.006133
Black Females			
Coefficient on CVD before age 60	$(1.251, 1.261)$	$(1.376, 1.381)$	$(1.502, 1.532)$
Coefficient on CVD after age 60	$(0.952, 1.016)$	$(0.974, 1.023)$	$(0.926, 0.975)$
Coefficient on Cancer before age 60	$(1.146, 1.171)$	$(1.185, 1.190)$	$(1.275, 1.280)$
Coefficient on Cancer after age 60	$(0.939, 1.018)$	$(0.955, 1.029)$	$(0.933, 1.012)$
Function Value	-0.001101	-0.004422	-0.006474

TABLE 7B: Marginal Identified Regions (break at age 65)

	1970–1980	1970–1990	1970–2000
White Males			
Coefficient on CVD before age 65	$(1.145, 1.150)$	$(1.338, 1.348)$	$(1.433, 1.444)$
Coefficient on CVD after age 65	$(0.925, 0.999)$	$(0.936, 1.015)$	$(0.913, 0.992)$
Coefficient on Cancer before age 65	$(1.081, 1.095)$	$(1.100, 1.100)$	$(1.183, 1.200)$
Coefficient on Cancer after age 65	$(0.923, 1.002)$	$(0.938, 1.017)$	$(0.936, 1.015)$
Function Value	-0.001786	-0.006560	-0.006840
White Females			
Coefficient on CVD before age 65	$(1.146, 1.146)$	$(1.201, 1.221)$	$(1.273, 1.273)$
Coefficient on CVD after age 65	$(0.932, 1.006)$	$(0.969, 1.043)$	$(0.958, 1.032)$
Coefficient on Cancer before age 65	$(1.091, 1.099)$	$(1.145, 1.149)$	$(1.232, 1.237)$
Coefficient on Cancer after age 65	$(0.938, 1.017)$	$(0.930, 1.004)$	$(0.956, 1.016)$
Function Value	-0.001405	-0.003768	-0.004637
Black Males			
Coefficient on CVD before age 65	$(1.145, 1.145)$	$(1.233, 1.249)$	$(1.388, 1.388)$
Coefficient on CVD after age 65	$(1.010, 1.024)$	$(0.931, 1.010)$	$(0.932, 1.011)$
Coefficient on Cancer before age 65	$(1.001, 1.031)$	$(1.001, 1.026)$	$(1.078, 1.098)$
Coefficient on Cancer after age 65	$(0.917, 0.996)$	$(0.934, 1.013)$	$(0.937, 1.016)$
Function Value	-0.000768	-0.006815	-0.010775
Black Females			
Coefficient on CVD before age 65	$(1.182, 1.192)$	$(1.254, 1.259)$	$(1.388, 1.388)$
Coefficient on CVD after age 65	$(0.932, 1.011)$	$(0.989, 1.033)$	$(0.882, 0.961)$
Coefficient on Cancer before age 65	$(1.102, 1.107)$	$(1.147, 1.147)$	$(1.169, 1.199)$
Coefficient on Cancer after age 65	$(0.931, 1.011)$	$(0.957, 1.016)$	$(0.987, 1.066)$
Function Value	-0.005226	-0.012324	-0.013312

TABLE 8: Bounds for Trends in cancer, excluding lung cancer

	1970–1980	1970–1990	1970–2000
White Males			
Independence	$(1.091, 1.093)$	$(1.126, 1.129)$	$(1.075, 1.076)$
No independence	$(1.091, 1.093)$	$(1.039, 1.045)$	$(1.236, 1.249)$
White Females			
Independence	$(1.091, 1.093)$	$(1.236, 1.250)$	$(1.334, 1.346)$
No independence	$(1.126, 1.129)$	$(1.239, 1.249)$	$(1.455, 1.458)$
Black Males			
Independence	$(1.084, 1.090)$	$(1.091, 1.093)$	$(1.091, 1.093)$
No independence	$(1.112, 1.115)$	$(1.201, 1.205)$	$(1.118, 1.119)$
Black Females			
Independence	$(1.059, 1.060)$	$(1.126, 1.129)$	$(1.239, 1.250)$
No independence	$(1.106, 1.111)$	$(1.143, 1.148)$	$(1.308, 1.319)$

TABLE 9: Marginal Identified Regions
(only Cancer multiplicative)

	1970–1980	1970–1990	1970–2000
White Males			
Coefficient on Cancer	$(0.520, 2.186)$	$(0.602, 2.124)$	$(0.654, 2.124)$
White Females			
Coefficient on Cancer	$(0.802, 1.610)$	$(0.890, 1.646)$	$(1.002, 1.698)$
Black Males			
Coefficient on Cancer	$(0.449, 2.356)$	$(0.484, 2.200)$	$(0.550, 2.332)$
Black Females			
Coefficient on Cancer	$(0.556, 2.284)$	$(0.644, 2.230)$	$(0.702, 2.332)$

TABLE 10: Marginal Identified Regions Starting at Age 40

	1970–1980	1970–1990	1970–2000
White Males			
Coefficient on CVD	$(1.126, 1.129)$	$(1.295, 1.296)$	$(1.389, 1.391)$
Coefficient on Cancer	$(1.001, 1.029)$	$(1.020, 1.035)$	$(1.134, 1.153)$
Function	-0.019	-0.036	-0.037
White Females			
Coefficient on CVD	$(1.091, 1.093)$	$(1.158, 1.160)$	$(1.236, 1.238)$
Coefficient on Cancer	$(1.091, 1.093)$	$(1.154, 1.160)$	$(1.201, 1.206)$
Function	-0.009	-0.016	-0.021
Black Males			
Coefficient on CVD	$(1.126, 1.129)$	$(1.201, 1.206)$	$(1.334, 1.346)$
Coefficient on Cancer	$(1.030, 1.034)$	$(1.063, 1.066)$	$(1.072, 1.074)$
Function	-0.005	-0.029	-0.039
Black Females			
Coefficient on CVD	$(1.158, 1.160)$	$(1.231, 1.235)$	$(1.334, 1.346)$
Coefficient on Cancer	$(1.096, 1.096)$	$(1.167, 1.172)$	$(1.160, 1.160)$
Function	-0.029	-0.035	-0.050