SEX DIFFERENCES IN MORBIDITY AND MORTALITY

Anne Case
Christina Paxson

Sex Differences in Morbidity and Mortality
Anne Case and Christina Paxson
NBER Working Paper No. 10653
August 2004
JEL No. I0, J1, J16

ABSTRACT

Women have worse self-rated health and more hospitalization episodes than men from early adolescence to late middle age, but are less likely to die at each age. We use 14 years of data from the U.S. National Health Interview Survey to examine this paradox. Our results indicate that the difference in self-assessed health between women and men can be entirely explained by differences in the distribution of the chronic conditions they face. Although on average women have worse self-rated health than men, women and men with the same chronic conditions have the same self-rated health. The results for hospital episodes are somewhat different. While the effect of poor health on hospital episodes is the same for men and women, men with respiratory cancer, cardiovascular disease, and bronchitis are more likely to experience hospital episodes than women who suffer from the same chronic conditions, implying that men may experience more severe forms of these conditions. The same is true for mortality. Although the effects of many chronic conditions on the probability of death are the same for women and men, men who report having cardiovascular disease and certain lung disorders are significantly more likely to die than women with these conditions. While some of the gender difference in mortality can be explained by differences in the distribution of chronic conditions, an equally large share can be attributed to the larger adverse effects of these conditions on male mortality. Is smoking the smoking gun? Conditions for which we find excess male hospitalizations and mortality are generally smoking-related.

Anne Case
367 Wallace Hall
Princeton University
Princeton, NJ 08544
and NBER
accase@princeton.edu

Christina Paxson
316 Wallace Hall
Princeton University
Princeton, NJ 08544
and NBER
cpaxson@princeton.edu

I. Introduction

Research on gender differences in health has brought to light an important paradox. Studies from a large number of countries find that women use more health services and report worse self-rated health than men. However, women are less likely to die than same-aged men throughout life, indicating that they may in fact be healthier.

There are several possible explanations for the worse self-rated health but lower mortality of women. There may be gender differences in the distributions of chronic conditions, driven by biological, behavioral or psychosocial factors (Verbrugge, 1989; Lawlor et al, 2001; Molarius and Janson, 2002). Women may be more likely than men to suffer from health conditions such as arthritis or headaches that result in poorer self-rated health but contribute relatively little to mortality risk; men may be more likely to have conditions such as cardiovascular disease or respiratory conditions that, although they contribute to worse self-rated health, also have relatively large effects on the probability of death. Although this explanation does not address *why* women and men have different distributions of health conditions, it can potentially account for the gender differences in self-rated health and mortality observed in the data.

Another possible explanation is that women are in fact healthier than men (as evidenced by their lower mortality), but simply *report* worse health on surveys. A commonly-held view is that women, being less stoical than men, are more likely to factor less serious ailments into their reports of poor health (Spiers et al, 2003). A twist on the same idea is that women are more accurate health reporters than men: they know more about their own health, perhaps due to their greater health care utilization, and are more willing to discuss their health and admit health problems to interviewers (Verbrugge 1989, Idler 2003). While these ideas are expressed in many papers, they are not uncontroversial. Some evidence contrary to this view comes from MacIntyre, Ford and Hunt (1999), who report on a study that asked men and women open-ended questions about health problems, followed by probes for specific health conditions. They find that men provided more complete information than women in response to the open-ended question. Although this evidence is not

paired with "objective" data on health that could be used to assess the accuracy of reports, it suggests that men are not less forthcoming than women in interviews.

A final possible explanation for the paradox is that the 'facts' underlying it are incorrect. Although the higher mortality of men is not in dispute, there is a literature that questions whether women have worse self-rated health and, more generally, higher rates of many measures of morbidity than men (Hunt and Annandale 1999). For example, MacIntyre, Hunt and Sweeting (1996) find evidence from two (relatively small) British samples that gender differences in reports of fair or poor health are significant only in very early adulthood, and that although women have higher rates of psychological distress than men, they do not have higher rates of many specific physical conditions. Some researchers find that gender differences in self-rated health become smaller at older ages (Case and Deaton 2003), and others find gender difference in self-rated health disappear in old age (Leinonen et al, 1997; Arber and Cooper, 1999).

The first explanation of the paradox — that gender differences in self-rated health and mortality are driven by differences across genders in the distributions of chronic conditions—is entirely consistent with women having higher rates of some illnesses than men, but lower rates of others. In fact, this first explanation *requires* that gender differences in prevalence rates vary across health conditions. Much of the literature that disputes whether women have higher rates of morbidity than men is concerned with establishing that women do not suffer from excess levels of *all* measures of ill health—a point with which we agree, and will provide evidence in support of below. We are concerned with the more precise issue of why women have worse self-rated health but lower mortality than men. As part of our investigation into gender differences in self-rated health and mortality, we also examine gender differences in hospital episodes. Unlike doctor's visits, which may be influenced by patient's demand for health care and perceptions of illness, hospital episodes provide a more "objective" measure of health status that can be compared with both self-rated health and mortality, and help provide a bridge between these two health measures.

Our analyses are based on fourteen years of U.S. data from the National Health Interview Survey (1986-1994, 1997-2001 NHIS) and its associated Multiple Cause of Death file, which provides information on deaths of those interviewed between 1986 and 1994. The surveys are ideal for our purpose because they contain information on self-rated health, hospital episodes and mortality, as well as detailed information on chronic health conditions that may contribute to both self-rated health and mortality. In addition, the survey covers large numbers of adult men and women: our analyses of self-rated health and hospitalization use a sample of 147,996 adult men and women surveyed between 1997 and 2001, and our analyses of mortality use a sample of 237,140 men and women ages 45-84 surveyed from 1986 to 1994. These large samples make it possible to obtain precise estimates of the gender differences in health measures at each age and of the effects of (sometimes rare) chronic health conditions on self-rated health and mortality.

Section II discusses the data and show that it is consistent with the basic facts of the puzzle: women have worse self-rated health and more hospitalization episodes than men from early adolescence to late middle age; but they are less likely to die at each age. Section III examines the validity of the explanations offered for this paradox. Our results indicate that the difference in self-assessed health between women and men can be entirely explained by differences in the distribution of conditions. Although on average women have worse self-rated health than men, women and men with the same sets of chronic conditions have the same self-rated health. The results for hospital episodes are somewhat different. While the effect of poor health on hospital episodes is the same for men and women, men with respiratory cancer, cardiovascular disease, and bronchitis are more likely to experience hospital episodes than women who suffer from the same chronic conditions, implying that men may experience more severe forms of these conditions. The same is true for mortality. Although the effects of many chronic conditions on the probability of death are the same for women and men, men who report having cardiovascular disease and certain lung disorders are significantly more likely to die than women with these conditions. While some of the gender difference in

3

mortality can be explained by differences in the distribution of chronic conditions, an equally large share can be attributed to the larger adverse effects of these conditions on male mortality.

Our results move us some distance in understanding the paradox between self-assessed health and mortality. That men and women with the same health conditions report the same self-rated health status, and the fact that poor health is equally predictive of hospitalization episodes for men and women, casts doubt on the idea that women and men use different standards for assessing self-rated health. However, our mortality results cannot be explained solely by differences in the distribution of chronic conditions. Men with smoking-related conditions are significantly more likely to die within two years than are women with the same conditions. These men will, on average, have had longer exposure to smoking in their lives, and their reports of (say) emphysema may indicate more advanced cases than do women's reports.

Our findings and their interpretation are important for several reasons. Self-reported health status is a tool often used to assess wellbeing. If women and men take different aspects of health into account, or weight them differently in their self assessments, it is important to understand why and how this is the case. In addition, if historically higher rates of smoking are responsible for the higher mortality rates men face in middle age, then we would anticipate the gap in age-adjusted mortality rates will close, given the changes observed in women's and men's smoking patterns over the last century. Finally, if the chronic conditions that lead to higher risk of mortality are given budget priority over those that lead to poorer self-assessed health, then research and treatment of important health conditions observed in women may be underfunded.

II. Data and Preliminary Evidence

A. <u>Data</u>

The data for this study are drawn from the National Health Interview Surveys (NHIS) conducted between 1986 and 1994, and 1997 to 2001, and from the associated NHIS Multiple Cause of Death Public Use Data File that contains information on deaths (as of 1997) of individuals surveyed

between 1986 and 1994. The NHIS is a cross-sectional household interview survey that covers the civilian non-institutionalized population living in the United States. The survey collects information on self-rated health, chronic health conditions, health care utilization, and socio-demographics.

Between 1986 and 1996, health information was collected for each member of sampled households. While all households were administered the same basic questionnaire, each household was randomly assigned to one of six "Condition Lists," and information was collected only on household members' experience with the chronic conditions included on the assigned list. The structure of the survey changed in important ways in 1997. While basic health information continued to be collected for all household members, information on chronic conditions among adults was collected only from a single "sample adult" in each household, who was asked about a full range of chronic conditions.

Because of the redesign and the timing of the Multiple Cause of Death File, we mainly rely on two distinct samples to study self-rated health and mortality. The 1997-2001 sample, consisting of 147,996 "sample adult" men and women ages 18 to 84, is used mainly for our analyses of self-rated health. We remove women who are either pregnant or have a child aged 1 or younger, in order to focus on non-pregnancy related health and hospitalizations. The fact that we have a complete set of information on chronic conditions on all sample adults is an advantage, since it permits us to more easily deal with co-morbidities across chronic conditions. The 1986-1994 sample is used primarily for our analyses of mortality. The sample includes men and women for whom vital status can be identified from the Multiple Cause of Death File. Because there are so few deaths among younger adults, we restrict our sample to 237,140 men and women ages 45-84 for whom vital status is known. Details on variable definitions are included in the Appendix.

B. Gender Differences in Self-Rated Health and Mortality

The NHIS data can be used to illustrate the paradox discussed in the last section. The two left-hand panels of Figure 1 present data on self-rated health status (SRHS), which in the NHIS is asked on a

5-point scale, where 1=Excellent, 2=Very Good, 3=Good, 4=Fair and 5=Poor. These figures use data on all household members from the combined 1986-2001 surveys, and show the average of SRHS, and the fraction of those who rate themselves in fair or poor health, by exact year of age for males and females.[1] These measures yield similar patterns of changes in health status with age. Boys have somewhat worse health status than girls prior to adolescence. However, the health status of girls becomes worse than that of boys around age 14. The gender gap in self-rated health status is greatest at age 20, and then slowly declines. By age 65 (using mean SRHS) or age 60 (using the indicator of fair or poor health), the female disadvantage in self-rated health status has vanished.

The top right panel shows the average number of hospitalization episodes over the previous 12 months, by age, for women and men. Between the ages of 20 and 60, women's number of hospitalizations is roughly constant, at 0.10 episodes a year. Only after age 60 do we observe the average number of episodes rising with age for women. In contrast, men's reports of hospitalizations rise with age from age 20. The pattern here mirrors what we observed for self-assessed health: women's excess hospitalizations is greatest at age 20. By late middle age, men's and women's reported hospitalizations equalize.

The bottom right panel of the figure shows the fraction of respondents in the 1986-1994 surveys (aged 25 to 74 at the time of the survey) who died within 24 months of their interviews.[2] This graphs show the well-known pattern of excess mortality among men at all ages. Although mortality is much lower at younger ages, the "excess" of male mortality is greatest at the youngest ages. The ratio of male to female mortality is fairly flat after age 40, increasing slightly at the oldest ages.

---

[1]These results are based on data for 1,602,650 individuals. Sample weights are used in the calculations. The health of those aged 16 and under is rated by an adult respondent. Children aged 17 have the option of reporting for themselves.

[2]This graph is based on 586,703 respondents of all ages for whom vital status could be determined.

Although women report worse health than men at all but the oldest ages, they are not uniformly more likely to report suffering from all chronic health conditions. Figure 2 shows prevalence rates for a selection of chronic health conditions for women and men by exact year of age. These chronic conditions were chosen to illustrate the diversity of patterns that appear in the data. Several of the chronic conditions, including frequent headaches, arthritis, and depression, are more prevalent for women at all ages. Others, such as asthma, display higher prevalence for women than men in middle age but not at the oldest ages. Yet others, such as reproductive cancers and cardiovascular disease, are more prevalent for women at younger ages but cross over to greater prevalence among men at older ages. Diabetes is equally prevalent for men and women at all ages. Emphysema and, to some extent, hypertension, are equally prevalent among younger men and women but more prevalent among men past age 60. These varied patterns confirm the view that women are not uniformly more likely than men to suffer from all types of ailments. However, they also suggest that women are more likely than men to suffer from conditions such as arthritis and frequent headaches that are not life-threatening but which could lead to poor self-rated health, and that men are more likely, at least at older ages when most deaths occur, to suffer from diseases such as cardiovascular disease and emphysema that are risk factors for mortality.

III. Chronic Conditions, Self-Rated Health and Mortality

A. <u>Self-Rated Health</u>

Methods

To assess routes into women's poorer health, we examine the prevalence and severity of chronic conditions collected in the 1997-2001 National Health Interview Surveys. We ruled out conditions that were benign enough to pose little health hazard (for example, hay fever), and restrict our attention to conditions for which a consistent definition was used over all years. We analyze 18 conditions that fall into six broad categories: pain (headache; neck/back ache; arthritis); respiratory conditions (bronchitis, emphysema, asthma, lung problems); circulatory conditions (cardiovascular

disease, diabetes, hypertension, circulatory problems); cancers (skin, stomach, reproductive, respiratory); chronic vision and hearing problems; and depression. Our 'condition' list is comprised of both diseases (for example, emphysema), and illnesses that may be symptoms of diseases (e.g. vision problems may be due to diseases of the eye, or a bi-product of diabetes.)

To quantify the extent to which sex differences in health status are due to differences in the presence of chronic conditions, or to differences in the impact of these conditions, or to unobserved factors, we need a method for decomposing health status. Initially, assume that only one chronic condition (C) exists, and that some subset of people suffer from this condition. (We extend this to multiple conditions in the empirical work below.) Then the probability of reporting fair or poor health (here referred to as "poor" health), can be decomposed into the probability of reporting poor health, conditional on the presence of the chronic condition, times the probability of observing the condition, plus the probability of reporting poor health, in the absence of the chronic condition, times the probability that the condition is absent:

$$P[Poor] = (P[Poor|C=1] \times P[C=1]) + (P[Poor|C=0] \times P[C=0])$$

Using this decomposition, and noting that the probability of the condition's absence is equal to 1 minus the probability of the condition's presence, we can characterize the difference in the probability that women and men report poor health, as follows:

$$
\begin{aligned}
P^W[Poor] - P^M[Poor] = \\
(P^M[Poor|C=1] - P^M[Poor|C=0]) \times (P^W[C=1] - P^M[C=1]) + \\
[(P^W[Poor|C=1] - P^W[Poor|C=0]) - (P^M[Poor|C=1] - P^M[Poor|C=0])] \times P^W[C=1] + \\
(P^W[Poor|C=0] - P^M[Poor|C=0])
\end{aligned}
\tag{1}
$$

This expression decomposes the gender difference in poor health into three components. The first quantifies the difference in poor health due to the difference in prevalence of the chronic condition between women and men. We will refer to this as the *prevalence effect*. The second term quantifies the difference in poor health due to the gender difference in the impact of the chronic condition on poor health. This we will refer to as the *severity effect*. The last term captures the difference in the probability of reporting poor health that is not due to the chronic condition.

This framework can be extended to multiple chronic conditions. Assume that the probability of reporting poor health depends linearly on indicators for the presence of a set of N chronic conditions (denoted as C_i, $i=1...N$) and indicator variables for age, race, and survey year, and a control for education (X):

$$P(Poor) = \sum_i^N \beta_i C_i + X\gamma + \epsilon ,\qquad (2)$$

We estimate equations of this form separately for men and women, and use the parameter estimates to compute prevalence and severity effects in the difference in poor health between women and men.

We estimate two variants of (2) that differ in their measures of chronic conditions. In the first variant, each of the N chronic condition indicators measures the presence or absence of one single ailment (e.g. arthritis, asthma, etc.) Individuals may have values of 1 for more than one of the 18 conditions we consider. This specification is parsimonious, but does not account for the fact that interactions between conditions may have additional effects on health status. In other words, the probability of poor health, given two chronic conditions, may not be the simple sum of the probability of poor health for each chronic condition observed individually:

$$P[Poor|C_1=1,C_2=1] \neq P[Poor|C_1=1,C_2=0] + P[Poor|C_1=0,C_2=1] .$$

In our NHIS data, 45% of respondents suffer from more than one chronic condition. The second variant allows for a large number of interactions between chronic conditions. In theory, we can incorporate the presence of multiple conditions by redefining our chronic conditions to measure each unique combination of ailments. In practice, it is not possible to allow for all possible combinations of the 18 ailments we follow. There are, for example, 153 different ways in which a person could suffer from exactly 2 of the 18 ailments, and 816 different ways in which a person could suffer from exactly 3 ailments. We can, however, capture the first-order interactions between conditions by defining "conditions" as single chronic ailments and all possible combinations of two chronic ailments. Individuals who suffer from cardiovascular disease and emphysema have a "CVD/emphysema" condition. To avoid double-counting, these individuals are not assigned the

"CVD only" or "emphysema only" conditions. We thus define 171 new conditions: 18 singletons plus 153 pairs of ailments (so that *N*=171). 73% of individuals report 2 or fewer ailments. For this group of respondents, the conditions we identify provide a complete accounting of the ailments that they face. For the remainder, the estimation yields an approximation for the effects of their conditions on health.

Results

We present a preliminary look at the extent to which the 18 ailments we follow affect men's and women's health status in Figure 3, which graphs the coefficients from ordinary least squares regressions as in equation (2), run separately for men and women. We regress an indicator variable for reporting fair or poor health (which we will refer to as 'poor' health) on all 18 chronic ailments, together with a complete set of age indicators, race indicators, survey year indicators, and a control for years of completed schooling. For both men and women, the condition that has the largest effect on self assessed health is stomach cancer, which is associated with an increased probability of reporting poor health of roughly 25 percentage points. Respiratory cancer and diabetes are each associated with an 18 percentage point increase in the probability of reporting poor health, for both men and women; depression and cardiovascular disease with a 15 percentage point increase; neck and back pain with a 5 percentage point increase. One condition that is not significantly associated with reported health status is skin cancer, which takes a small, negative insignificant coefficient for both men and women. Skin cancer is the only condition on our list that is not significantly associated with poor health.

The solid line in Figure 3 marks the 45 degree line: if the increase in the probability of reporting poor health in the presence of a given chronic condition were equal for men and women, then the regression coefficients reported for the two sexes would fall on this line. Figure 3 shows remarkable similarity between the sexes in the association between poor health and the presence of each chronic ailment. Tests for differences between men and women in the 18 ailment coefficients

10

were significant for only 6: headaches, back and neck pain, arthritis, hypertension, depression, and lung problems. Of these six, only the difference associated with lung problems is large in magnitude, where the probability of poor health is 7.0 percentage points higher for men than for women. Lung disorders, and more generally chronic illnesses associated with smoking, will be seen to affect hospitalizations and mortality in what follows.

The near equality between men and women in their health responses to chronic conditions suggests that gender differences in response to chronic conditions is unlikely to lead to an explanation of gender differences in reported health status. Indeed, in the formal decomposition presented below, we find no role for the severity effect in explaining the self-assessed health differentials.

Table 1 presents prevalence rates for chronic conditions, and the differences in the prevalence rates between women and men for the 18 ailments we use in our analysis. The last column shows the age- and race-adjusted excess percentage point prevalence among women for each ailment, above that expected for men. These were estimated by regressing each ailment on a complete set of age, race and survey year indicators, a measure of completed education, and an indicator that the respondent is female, for 147,996 individuals surveyed between 1997 and 2001. An asterisk (*) denotes those conditions for which the prevalence rates are significantly different between women and men at a 5% level. Women have significantly higher rates of pain (headache, other pain, arthritis), and some respiratory conditions (bronchitis, asthma, lung problems other than cancer). They are also significantly more likely to suffer from reproductive cancers, hypertension, vision problems and depression. Men are significantly more likely to suffer from smoking-related ailments (emphysema, respiratory cancer), some circulatory problems (cardiovascular disease, diabetes), and hearing loss. Men are significantly more likely to have two-ailment conditions that involve emphysema, respiratory cancer, diabetes, and cardiovascular disease, while women are significantly more likely to have two-ailment conditions that involve pain, asthma, bronchitis, vision, and depression. The prevalence rates for the set of 171 chronic conditions and condition

11

interactions are presented in Figure 4, where it is clear that the difference in prevalence rates between women and men are larger for conditions that women are more likely to report.

To estimate the severity and prevalence effects of these conditions, we first regress an indicator for reporting fair or poor health on a vector of chronic conditions, as in equation (2) above, in separate regressions for men and women. Differences in men's and women's chronic condition coefficients, multiplied by the average prevalence rate of each condition (denoted as $\overline{C_i}$), form the basis of our severity effect:

$$\textit{severity effect} = \sum_i (\beta_i^W - \beta_i^M)\,\overline{C_i}\,.$$

Differences in these prevalence rates between men and women, multiplied by the condition's β (averaged between the β estimated for men and women), form the basis of the prevalence effect:

$$\textit{prevalence effect} = \sum_i (\overline{C_i^W} - \overline{C_i^M})\,\overline{\beta_i}\,.$$

The residual difference in women's and men's self-assessed health—that unexplained by the severity and prevalence effects above—combines differences produced by gender differences in average values of X (age, education and race), gender differences in the effects of these variables on health (e.g. in γ), and unexplained differences.

Our estimates of these effects are presented in Table 2. The first column presents results from regressions that include 18 chronic conditions with no interactions between the conditions. The second column presents results from regressions that included 171 chronic conditions and chronic condition interactions. On average, 14.3% of women aged 18 to 84 reported fair or poor health. The likelihood of a woman reporting poor health is 2.5 percentage points greater than is that for a man. With or without the chronic condition interactions, we find that the difference between women's and men's health is fully explained by differences in the prevalence of conditions they face. From our analysis that includes condition interactions (column 2), we predict women to be 2.3 percentage points more likely to report poor health based on the conditions they report. There is little role for the severity effect: we find men are 0.13 percentage points more likely to report poor health than are women with the same chronic conditions.

In tests run on the health status regressions, we found that the condition interaction variables were jointly significant. However, their absence (column 1) makes little difference to the health status decomposition.

B. <u>Hospitalizations</u>

A different look at health is provided by the number of hospitalization days and episodes men and women report having had in the past 12 months. (Hospitalization refers only to in-patient stays.) Hospitalization may provide a more objective measure of health status than do self-reports. Unlike doctors visits, which patients can generally schedule themselves, hospital admission is controlled by health providers, who may fail to admit patients with less serious health problems.

If it were the case that women were less stoical than men, and more likely to factor minor ailments into their reports of poor health, then we would expect women's hospitalizations to be less responsive to their reports of poor health than are men's. Instead, we find equally large responses between hospitalization episodes and reports of poor health for men and women. In regressions of the number of reported hospitalization episodes on an indicator of reporting fair or poor health, with a complete set of indicators for age, race, survey year and a measure of educational attainment, we find:

Women: *Number of episodes = 0.341 Fair/Poor Health*
$$(0.006)$$

Men: *Number of episodes = 0.338 Fair/Poor Health*
$$(0.007)$$

For both women and men, a subjective report of fair or poor health translates on average into one-third extra hospitalization episode a year.

Women on average have 0.034 more hospitalization episodes per year than do men, and 0.052 extra hospital days. As observed in Figure 1, these differences are largest in young adulthood.

Using the same methods presented for health status above, we can decompose hospitalization days and episodes into severity, prevalence and residual effects. We present these decompositions in Table 3, where the first three columns present results for hospitalization episodes, and the second three for hospitalization days. For both episodes and days, we present results for our 18 chronic conditions without condition interactions (columns 1 and 4), and with interactions (columns 2 and 5). Consistent with our findings for self-reported health status, the difference in the distribution of chronic conditions goes some way toward explaining women's greater hospitalization. We find a prevalence effect for hospitalization episodes of 0.0176, in our analysis containing the 171 conditions and condition interactions. This is half of the total difference we observe in hospitalization episodes between men and women. We find a prevalence effect for hospital days of 0.0941, which is almost twice as large as the total gap between women and men's hospitalization days. Based on the distribution of conditions alone, we would expect women to spend even more days in the hospital, relative to men, than we observe in the data.

These prevalence effects are offset by differences in the impact of chronic conditions on hospitalizations for men and women. We find, for both episodes and days, that when men and women report the same chronic conditions, men have more episodes and more hospital days than do women. The conditions responsible for this severity effect can be seen in Figure 5, which presents coefficients from regressions of men and women's hospitalization episodes on 18 chronic conditions. Men suffering from respiratory cancer, stomach cancer, bronchitis, and cardiovascular disease report significantly greater numbers of hospital episodes than do women reporting the same conditions.

C. Mortality

Methods

The last two sections provided evidence that the worse self-rated health of women can be entirely explained by gender differences in the distribution of chronic health conditions, but that men are

more likely than women with the same health conditions to have hospitalizations. This section turns to mortality. We study whether men and women who report the same health conditions are equally likely to die in the two years following the survey, or whether men have excess mortality given health conditions. We also decompose gender differences in mortality into prevalence and severity effects, as was done for self-rated health and hospitalizations.

As in the last sections, we start by specifying mortality as a function of health conditions (C_i) and other sociodemographic factors (X):

$$P(D) = \sum_i \beta_i C_i + X\delta + \epsilon = C\beta + X\gamma + \epsilon,$$

(3)

where D is indicator that the respondent died within 24 months of the survey. These equations are estimated separately for women and men. Because there are few deaths among younger adults, the samples include only those aged 45 to 84. Our primary focus is on whether the effects of chronic conditions on mortality differ for women and men.

Although the framework is identical to that described in the last sections, data issues require that we use different and somewhat more complex estimation methods. The NHIS mortality information is available only for those surveyed before the 1997 redesign. As discussed in the data section, the NHIS followed very different procedures for collecting information on health conditions before 1997: instead of asking a subset of adults about all health conditions, the survey asked all adults about a subset of health conditions on the condition list to which they were randomly assigned. As a result, we have only incomplete information on the variables in C for any one individual. While this method of collecting information allows for the calculation of accurate prevalence rates with minimal respondent burden, it complicates analyses that require information on co-morbidities. Consider, for example, the strategy of estimating equation (3) one "condition list" at a time, so that we regress the death indicator on the conditions included in Condition List 1 using the sample that was assigned to this list, and repeat this for each of the condition lists. The estimates of the effects of the conditions on the probability of mortality (i.e. the estimates of β) will be biased, since they will reflect the effects of conditions that are unmeasured but correlated with

15

the conditions that are measured. For example, if those who have heart disease also have an elevated risk of emphysema (perhaps because smoking increases the risk for both diseases), the estimates of the effects of both heart disease and emphysema, which appear in different condition lists, are likely to be biased upward.

The bias due to incomplete information on chronic conditions can be corrected with supplementary information on the covariance of each pair of chronic diseases that appear in different condition lists. Our strategy is to use information on these covariances from the 1997-2001 NHIS surveys to correct for bias due to missing information on co-morbidities.[3] To see how these biases can be corrected, consider a simplified version of (3) in which the controls for the sociodemographic variables have been suppressed:

$$D = C_1 \beta_1 + C_2 \beta_2 + ... + C_M \beta_M + \epsilon = C\beta + \epsilon, \tag{4}$$

where C_j represents an $N x k_j$ matrix of indicators for whether individuals have chronic conditions included in condition list j, and there are a total of M condition lists. (In practice, we include controls for sociodemographic variables. The methods described below are easily extended to the case in which other control variables are included.) Note that the 1986-1994 data, which we refer to as the "censored" sample, contain information on D but incomplete information on C. The 1997-2001 data, which we refer to as the "supplemental" sample, has complete information on C but no information on D.

The equation for bias correction can be obtained by first defining a matrix Z, which is equal to C but with unobserved values of conditions set to zero, and considering the ordinary least squares estimates of β that result when Z is used in place of C:

$$\tilde{\beta} = (Z'Z)^{-1}(Z'C)\beta + (Z'Z)^{-1}Z'\epsilon \tag{5}$$

and

$$plim \ \tilde{\beta} \ = \ \Sigma_{ZZ}^{-1} \Sigma_{ZC} \beta \tag{6}$$

---

[3]The strategy of using supplemental data in this way has been used to correct for biases in the estimates of the effects of lead on IQ (Marais and Wecker, 1998).

where Σ_{ZZ} is the variance-covariance matrix of the censored data matrix Z and Σ_{ZC} is the covariance matrix of Z and C. Equation (6) indicates that $\tilde{\beta}$ is an inconsistent estimator of β, and suggests the following corrected estimator (denoted $\hat{\beta}$):

$$\hat{\beta} = \hat{\Sigma}_{ZC}^{-1} \hat{\Sigma}_{ZZ} \tilde{\beta} \tag{7}$$

where $\hat{\Sigma}_{ZZ}$ and $\hat{\Sigma}_{ZC}$ are consistent estimators of the covariance matrices. The former can be obtained from the censored sample. The matrix Σ_{ZC} contains cross-products for pairs of conditions that are from the same condition lists (and so can be estimated from the censored sample alone) and for pairs of conditions that appear in different condition lists (and so must be estimated from the supplemental sample.) Details on how these covariance matrices were estimated, and on estimates of the standard errors for $\hat{\beta}$, are in the Appendix.

Although this estimation method delivers consistent estimates of β for women and men, the lack of complete information on all conditions is restrictive. We cannot, for example, estimate models similar to those shown for self-rated health and hospitalizations, in which interactions between conditions are included. We also cannot use non-linear functional forms for (4)—for example, logit or probit models—since the bias correction formula are only correct for the linear model. Finally, we must restrict the list of health conditions to those that are measured in similar ways in the earlier censored sample and the later supplemental sample. We examine fourteen chronic conditions, including arthritis and skin cancer (from Condition List 1); digestive cancer (List 3); diabetes, frequent headaches and reproductive cancer (List 4); CVD, hypertension, and other circulation problems (List 5); and bronchitis, asthma, emphysema, other (non-cancerous) lung disorders, and respiratory cancer (List 6).

Results

Table 4 shows results from two sets of estimates of the effects of chronic conditions on the probability of death within 24 months of the survey. The first two columns show estimates of very

simple variants of (3), in which the indicator for death is regressed on only *one* of the chronic conditions indicators, plus a set of indicators for age and race, and a control for years of education. Thus, each cell of each row shows the marginal effects of the condition on the probability of death for women (row 1) and men (row 2), controlling for no other conditions. The third column provides the p-value from a test that the coefficients for women and men are identical. These simple results are shown so that they can be contrasted with those in the last three columns, which contain results from the complete model that includes the bias correction discussed above.

The results with and without bias correction are generally consistent with those shown above for self-rated health and hospitalizations. The chronic conditions that lead to the largest increases in the probability of death—CVD, emphysema, other lung problems, respiratory cancer, digestive and reproductive cancers and diabetes—also elevate reports of poor health and hospitalizations. Conditions that have smaller or even no effects on mortality, such as headaches or hypertension, have smaller (although significant) adverse effects on self-rated health.

Estimates of the marginal effects of conditions on mortality are generally smaller when the complete set of conditions is included and the bias correction is used. For example, cardiovascular disease (CVD) is estimated to increase the probability of death by 2.9 percentage points for women and 4.6 percentage points for men when no conditions other than CVD are included. In the bias-corrected model, these effects decline to 2.1 (for women) and 3.6 (for men) percentage points. These declines reflect the fact that individuals often have multiple conditions, something illustrated in the results in the last section.

Correcting the bias tends to reduce the men's coefficient by more than the women's coefficients. For example, without adjustment for co-morbidities, the difference in the effects of CVD on mortality is –1.7 percentage points (2.9 for women minus 4.6 for men). The adjustments for co-morbidities reduces this difference to –1.5 (2.1 for women minus 3.6 for men). The pattern of greater reductions in men's than women's coefficients is evident for 10 of the 14 conditions. This is consistent with the evidence on patterns of co-morbidities discussed in Section A, above, which

noted that men are more likely than women to have combinations of more life-threatening conditions (e.g. emphysema, respiratory cancer, diabetes, and cardiovascular disease), whereas women's co-morbidities are clustered around conditions that are less likely to be fatal (e.g. conditions that involve pain, asthma, and bronchitis).

Even after adjusting for co-morbidities, the presence of health conditions is often associated with larger increases in mortality for men than for women. This can be seen from the last three columns in Table 4, or in the graphs of these coefficients in Figure 6. (The top panel of Figure 6 shows all fourteen conditions. The bottom panel excludes respiratory and digestive cancer, which have very large effects of the probability of death.) Of the eight conditions that have marginal effects on the probability of death in excess of 0.01, five have larger effects on male mortality (these include stomach cancer, CVD, emphysema, other lung disorders, and respiratory cancer) and three have larger effects of female mortality (diabetes, reproductive cancer, and bronchitis). Although the majority of these differences are not statistically significant (see last column of Table 4), the results taken together suggest that serious health conditions pose a greater threat to men than women.[4]

This can be seen more formally by decomposing the difference between female and male rates into prevalence and severity effects, as was done for self-rated health and hospitalizations in the last two sections. In this sample of 45-84 year olds, 2.6 percent of women and 4.2 percent of men die within two years of the survey, resulting in a difference between women and men of –1.6 percent. This –1.6 percent difference can be decomposed into a prevalence effect of –0.2 percent, a severity effect of –0.6 percent, and a remainder unexplained by health conditions of –0.8 percent. Thus, of the 50% of the gender difference in mortality that can be explained by these 14 chronic health conditions, 25% (–0.02/–0.08) can be explained by men being more likely to suffer from the

---

[4]Consistent with our finding for respiratory cancer, other research indicates that being female is associated with longer survival after surgery and chemotherapy treatments for lung cancer (Patel et al, 2004).

chronic conditions that have larger effects of mortality, and 75% (–0.06/–0.08) is explained by men having a greater probability of dying than women with the same chronic conditions. However, this decomposition provides a misleading estimate of the size of the severity effect. Fully half of the severity effect is due to the gender difference in the estimated effect of arthritis on mortality. As shown in Table 4, arthritis is not associated with male mortality and is *negatively* associated with female mortality, so that it contributes to the severity effect for men. In addition, since arthritis is very prevalent, it is heavily weighted in the calculation of the severity effect. When the effects of arthritis on mortality are set to zero for both women and men—which seems sensible, given that we do not think that arthritis really protects women against death—the severity effect drops to –0.2 percent, equal to the prevalence effect.

Is smoking the smoking gun?

Cardiovascular disease is the major contributor to the severity effect (because of its high prevalence combined with the gender difference in its effect on mortality). Stomach cancer, emphysema, other lung disorders and respiratory cancer are also contributing factors. All of these conditions are associated with smoking.[5] It is possible that the higher smoking rates of men result in more severe forms of these conditions when they occur. The 1986-1994 NHIS surveys do not contain information on the smoking history of respondents, so we cannot examine directly whether the effects of these conditions on mortality are higher among smokers. We can, however, use the 1997-2001 NHIS to document that men do have substantially greater life-time exposure to smoking than women. Figure 7 shows the average number of years of smoking for men and women, by age at the

---

[5]The link between smoking and stomach cancer is not as well known as the association of smoking with the other four conditions. However, the American Cancer Society reports that "Smoking increases stomach cancer risk, particularly for cancers of the proximal stomach (the upper portion of the stomach closest to the esophagus). The rate of stomach cancer is approximately doubled in smokers."
http://www.cancer.org/docroot/CRI/content/CRI_2_4_2X_What_are_the_risk_factors_for_stomach_cancer_40.asp

time of the survey, for sample adults in these surveys.[6] Here we see stark differences in lifetime smoking between men and women. Even including those who never smoked, 75-year-old men surveyed between 1997 and 2001 report an average of more than 25 years of smoking, in contrast to fewer than fifteen years of smoking for women. Years of smoking are lower among those in younger cohorts so that, among 45-year-olds, men have on average only 2 more years of smoking over their lives than women. If gender differences in life-time tobacco use are generating excess male mortality, then smaller gender gap in smoking among these younger cohorts may reduce excess male mortality *both* through the prevalence effect and the severity effect.

This explanation for the differences in the mortality effects that we find is consistent with research by Pampel (1997), who concludes that smoking fully explains the recent narrowing in mortality differentials observed between men and women. Pampel cites Valkonen and van Poppel (1997) as having concluded that 40% of total sex differences in life expectancy at age 35 could be explained by smoking. If Valkonen and van Poppel are correct, then combining the smoking results with differences in the distributions of non-smoking related chronic conditions may bring us most of the distance needed to explain excess male mortality.

IV. Discussion and Conclusions

The results presented above go partway to resolving the paradox with which we started, but do not resolve it entirely. The hypothesis that gender differences in self-rated health can be entirely explained by gender differences in the distribution of conditions is confirmed. The hypothesis that women and men form assessments of their health in different ways is not consistent with either the evidence that women and men with the same health conditions are equally likely to report being in

---

[6]For those who never smoked, years of smoking is set to zero. For those who report ever smoking, years of smoking is calculated as the year the respondent quit (or the current year if she or he still smokes) minus the year in which he or she began smoking. This, years of smoking will be overstated for those who do not smoke constantly between the start date and their most recent quit date (or the date of the interview.)

poor health, or with our findings that reports of poor health are equally predictive of hospitalization episodes for men and women. However, the evidence for hospitalizations and mortality indicate that men with some health conditions are more likely to be hospitalized and to die than women with the same conditions. The "severity effects" that disadvantage men are driven mainly by a small number of smoking-related conditions: CVD (for hospital episodes), and CVD, emphysema, and other lung disorders (for mortality).[7] The remaining question is why, if these conditions are more likely to result in hospitalizations and death for men, do they not also result in greater reports of poor health for men?

One partial answer to this question is that some of these conditions do produce both excess male mortality and higher reports of poor health for men relative to women. The results in Figure 3 indicate that men with emphysema and "other lung" problems are more likely than women to report poor health. Although gender differences in these coefficients are statistically significant only for lung problems and not for emphysema, they point to the possibility that men experience more severe forms of these conditions—something that is consistent with the evidence of greater rates of smoking among men presented above. However, the finding that men with CVD are equally likely to report poor health but more likely to experience hospitalizations and die than women with CVD is more difficult to explain.

Another possible answer is that men tend to report health conditions only when they are at more severe or advanced stages. For example, a woman with mild angina might report having CVD, whereas a man might not report CVD until he has experienced a heart attack. This would explain why reported CVD has a larger effect on male than female mortality. However, to reconcile the results for self-rated health with those for mortality, it would also have to be the case that women with milder forms of CVD are as likely to report poor health as men with more severe forms of

---

[7]Stomach and respiratory cancer also have larger effects on mortality and hospitalizations for men than for women. However, because these conditions have very low prevalence rates, they contribute little to the overall severity effects for mortality or hospitalizations.

CVD. In other words, women would have to be more likely than men to report less severe forms of conditions *and* to factor milder forms of conditions into their reports of poor health.

Although logically possible, we do not think this answer is plausible. For many health conditions we examine, there are no gender differences in the effects on both self-rated health and mortality. This is true even for conditions such as diabetes and bronchitis, that can display varying degrees of severity. If women systematically reported milder forms of all conditions, we would expect to see these conditions having smaller effects on women's mortality. Additional evidence comes from looking at samples of individuals who died from specific causes (e.g. acute myocardial infarction, stroke) within two years of the survey, and examining whether those individuals reported conditions associated with those causes of death. If men are more likely than women to under-report conditions, then men who died should be less likely than women to have reported having the condition that resulted in the death. The results in Table 5 indicate that this is not the case. Women and men whose causes of death were digestive cancer, diabetes, heart attack, stroke, hypertensive disease, and respiratory cancer were equally likely to have reported the relevant condition at the preceding interview. Men who died of respiratory disease (bronchitis, asthma, emphysema or chronic obstructive pulmonary disease) were more likely than women to have reported one or more of the conditions leading up to this cause of death. It should be noted that reports of the relevant conditions are often low for both men and women, indicating that there may be under-reporting of chronic conditions, due either to unwillingness to disclose health conditions or lack of knowledge about those conditions.[8] However, we find no evidence that men report conditions less than women.

A final explanation is that, for at least some conditions, the symptoms individuals experience may convey little information about the severity of the disease. This may be especially relevant for CVD, several components of which can be symptomless. It is possible that women and men who report having CVD experience similar symptoms and therefore report (on average) the

---

[8]Additional evidence on under-reporting of health conditions in Canada can be found in Baker, Stabile and Deri (2001).

same self-rated health, when in fact the men's disease is more severe (perhaps due to their greater levels of smoking). This explanation is consistent with our finding that men and women with CVD are equally likely to report poor health, that men with CVD are only slightly more likely than women to experience hospital episodes, and that men with CVD are more than 1.5 times more likely to die than women who report CVD.

Additional evidence on this point comes from the 1988-1994 National Health Examination Study (NHANES III), which collected health information using both interviews and medical exams for a nationally representative sample.[9] We use NHANES III data for a sample of men and women ages 30-74 to compare measures of cardiovascular health for men and women who report having cardiovascular disease.[10] Specifically, we examine whether men who self-report CVD have had more heart attacks, are more likely to report episodes of severe chest pain, are more likely to smoke, and have higher blood pressure, triglycerides, and cholesterol than do women who report CVD. The first three of these measures—heart attacks, chest pain and smoking—are self-reported and are certainly not "symptomless". However, measures of blood pressure, triglycerides and cholesterol are obtained from the examination and laboratory tests. These cannot be subject to gender-specific reporting bias, and respondents do not know the results of their examinations and blood work when they report whether they have cardiovascular disease.

We also compare men's and women's values of the Framingham risk score, which provides a useful summary measure of the propensity to develop coronary heart disease (CHD), defined as defined as angina pectoris, myocardial infarction, coronary insufficiency, and coronary heart disease death, over a 10-year period. This scoring method was developed by researchers using data from the

---

[9]Additional information on the NHANES III can be found at the NCHS website: http://www.cdc.gov/nchs/products/elec_prods/subject/nhanes3.htm

[10]This age range is used since the Framingham risk score, discussed below, was developed only for individuals aged 30-74.

Framingham study, and is described in detail in Wilson et al (1998).[11] The ingredients that go into

calculating the score include age, current smoking, diabetes, total cholesterol, HDL cholesterol, and

blood pressure. It is important to note that method of scoring CHD risk varies by gender, and that

men generally receive higher (worse) scores than women with the same profiles of risk factors. For

example, a man with very low HDL (below 35 mg/dl) is assigned more points on the scale than a

women with very low HDL, indicating that men from the Framingham sample with low HDL were

(all else equal) more likely to develop CHD than women. The Framingham result, that men are

more likely to develop CHD than women with the same risk factors, is consistent with our finding

that men who report cardiovascular disease are more likely to die than women with the same

conditions.

Table 6 shows that men who self-report cardiovascular disease at baseline are more likely to

experience some, but not all, risk factors. They are significantly more likely to currently smoke, and

have experienced more heart attacks on average than women who also report cardiovascular

disease. Men are 6 percentage points more likely than women to report severe chest pain, although

this difference is only marginally significant. Men are also significantly more likely to have

diastolic blood pressure and HDL cholesterol in the "high risk" range. Women, however, are

significantly more likely to have high total cholesterol. The Framingham risk score, which

summarizes the combined effects of blood pressure, total cholesterol, and HDL cholesterol, as well

as smoking and diabetes, is nearly twice as high for men as for women who self-report

cardiovascular disease. The Framingham 10-year risk of developing CHD for men is 19.6 percent,

---

[11]The scoring system is based on age-adjusted Cox proportional hazard models, estimated
by gender, that relate the development of coronary heart disease (defined as angina pectoris,
myocardial infarction, coronary insufficiency, and coronary heart disease death) within a 12-year
follow-up period to baseline measures of total cholesterol, HDL cholesterol, diastolic blood
pressure, diabetes and smoking. The raw score in then scaled so that it measures the 10-year risk
of experiencing CHD given the individual's baseline characteristics.

and for women is only 10.7 percent.[12] These results indicate that, among men and women who

report cardiovascular disease, men generally have worse risk profiles than do women, which may

account for their higher mortality rates.

In summary, our results indicate that women and men who experience the same health

conditions are equally likely to report being in poor health. And, many health conditions have

effects on hospitalizations and mortality that are similar for men and women. However, men with

some health conditions—notably those associated with smoking—are more likely than women with

these conditions to experience hospitalizations and death.   Overall, the results suggest that the

paradox of worse self-rated health for women and higher mortality for men has a fairly

straightforward explanation that does not rest on systematic differences between women and men in

how health is reported.

---

[12]It should be noted that the 10-year risks are likely to be higher for this sample of men
and women who report having CVD. The Framingham method assigns raw scores to risks for
individuals who have not yet been diagnosed with CHD at baseline. In addition, the Framingham
risks are based on samples of white individuals from a New England suburb. However, the risk
scores provide a useful method of aggregating across the different risk factors for heart disease,
ven if they do not assign accurate risks to individuals in the NHANES III sample.

References

Arber, S. and Cooper, H. (1999). Gender differences in health in later life: The new paradox? *Social Science and Medicine* 48. 61-76.

Baker, M., Stabile, M. and Deri, C. (2001). What do self-reported, objective measures of health measure? NBER Working Paper 8419, *Journal of Human Resources*, forthcoming.

Case, A. and Deaton, A. (2003). Broken down by work and sex: How our health declines. NBER Working Paper 9821.

Hunt, K. and Annandale. E. (1999). Relocating gender and morbidity: Examining men's and women's health in contemporary Western societies. *Social Science and Medicine* 48. 1-5.

Idler, E. L. (2003). Discussion: Gender differences in self-rated health, in mortality, and in the relationship between the two. *The Gerontologist* 43(3). 372-75.

Lawlor, D.H., Ebrahim, S. and Davey Smith G. (2001). Sex matters: Secular and geographical trends in sex differences in coronary heart disease mortality. *British Medical Journal* 323. 541-45.

Leinonen, R., Heikkinen, E. and Jylha, M.. (1997). Self-rated health and self-assessed change in health in elderly men and women — a five-year longitudinal study. *Social Science and Medicine* 46(4-5). 591-97.

MacIntyre, S., Hunt, K., and Sweeting, H.. (1996). Gender differences in health: Are things really as simple as they seem? *Social Science and Medicine* 42(4). 617-24.

Marais, L.M. and Wecker, W.E. (1998). Correcting for omitted-variables and measurement-error bias in regression with an application to the effect of lead in IQ. *Journal of the American Statistical Association* 93(442). 494-505.

Molarius, A. and Janson, S. (2002). Self-rated health, chronic diseases, and symptoms among middle-aged and elderly men and women. *Journal of Clinical Epidemiology* 55. 364-70.

Pampel, F. C. (2002). Cigarette use and the narrowing sex differentials in mortality. *Population and Development Review* 28(1). 77-104.

Patel, J.D., Bach, P.B. and Kris, M.G. (2004). Lung cancer in women: A contemporary epidemic. *Journal of the American Medical Association* 291(14): 1763-1768.

Spiers, N., Jagger, C., Clarke, M. and Arthur, A. (2003). Are gender differences in the relationship between self-rated health and mortality enduring? Results from three birth cohorts in Melton Mowbray, United Kingdom. *The Gerontologist* 43(3). 406-11.

Valkonen, T. and van Poppel, F. (1997). The contribution of smoking to sex differences in life expectancy: Four Nordic countries and The Netherlands 1970-1989. *European Journal of Public Health* 7. 302-10.

Verbrugge, L. M. (1989). The twain meet: Empirical explanations of sex differences in health and mortality. *Journal of Health and Social Behavior* 30. 282-304.

Wilson, P. W.F., D'Agostino, R.B., Levy, D., Belanger, A.M., Silbershatz, H., and Kannel, W.B., (1998). Prediction of coronary heart disease using risk factor categories. *Circulation* 97(18). 1837-47.

Table 1. Chronic condition prevalence rates

Condition	Prevalence Women	Prevalence Men	Excess prevalence in women (percentage points)
Headache	24.0	11.4	13.2*
Other pain	39.5	35.6	3.7*
Arthritis	28.0	19.2	7.2*
Bronchitis	6.7	3.2	3.3*
Emphysema	1.4	2.0	−0.8*
Lung problems	2.0	1.4	0.5*
Asthma	10.8	7.8	3.1*
Diabetes	6.6	6.1	−0.3*
Circulatory problems	0.5	0.4	0.05
Cardiovascular disease	13.3	12.9	−1.1*
Hypertension	26.5	23.3	0.5*
Skin cancer	1.9	2.3	−0.5*
Stomach cancer	0.1	0.1	−0.01
Reproductive cancer	3.3	1.8	1.2*
Respiratory cancer	0.2	0.4	−0.2*
Vision	11.2	8.3	2.2*
Hearing	3.7	6.0	−3.0*
Depression	12.8	9.7	2.9*
Number of observations	81704	66292	147996

Notes: Excess prevalence coefficients reported in column 3 are the coefficients on an indicator that the respondent is female, in OLS regressions of each condition on a complete set of age, survey year and race indicators, a variable for completed education, and a sex indicator. An asterisk (*) denotes that the difference in the rates for men and women are significant at the 5% level.

Table 2. The decomposition of health status

	The fraction of respondents reporting Fair/Poor health	
	Allowing for 18 Conditions	Allowing for 18 Conditions + 153 Condition Interactions
Women	0.1428	0.1428
Men	0.1181	0.1181
Difference (women–men)	0.0247	0.0247
Severity effect	0.0065	–0.0013
Prevalence effect	0.0261	0.0233
Difference (women–men) for those with no chronic conditions	–0.0079	0.0028
F-test for the joint significance of 153 condition interaction terms (*p*-value) for women's regressions		8.07 (0.0000)

Table 3. The decomposition of hospitalization episodes

	Dependent Variable:					
	Hospitalization Episodes			Hospitalization Days		
	18 Conditions	18 Conditions + 153 Condition Interactions		18 Conditions	18 Conditions + 153 Condition Interactions	
	All ages	All ages	Ages 45-84	All ages	All ages	Ages 45-84
Women	0.1502	0.1502	0.1870	0.7286	0.7286	1.019
Men	0.1160	0.1160	0.1853	0.6769	0.6769	1.094
Difference (women–men)	0.0343	0.0343	0.0016	0.0517	0.0517	–0.0749
Severity effect	–0.0133	–0.0082	–0.0030	–0.1170	–0.0861	–0.0800
Prevalence effect	0.0196	0.0176	0.0089	0.0980	0.0941	0.0268
Difference (women–men) for those with no chronic conditions	0.0252	0.0241	–0.0045	0.0642	0.0362	–0.0281

Table 4: Marginal effects of conditions on 2-year mortality

Condition	unadjusted for co-morbidities			adjusted for co-morbidities		
	women	men	p-value: women=men	women	men	p-value: women=men
arthritis	−0.0032 (0.0023)	0.0067 (0.0028)	0.006	−0.0099 (0.0029)	0.0018 (0.0039)	0.016
skin cancer	0.0006 (0.0084)	−0.0055 (0.0069)	0.577	−0.0050 (0.0077)	−0.0083 (0.0079)	0.762
digestive cancer	0.1630 (0.0189)	0.1940 (0.0199)	0.259	0.1571 (0.0169)	0.1973 (0.0219)	0.146
diabetes	0.0401 (0.0042)	0.0415 (0.0047)	0.826	0.0355 (0.0043)	0.0305 (0.0058)	0.489
frequent headache	−0.0015 (0.0036)	−0.0037 (0.0057)	0.746	−0.0054 (0.0037)	−0.0180 (0.0068)	0.100
reproductive cancer	0.0463 (0.0082)	0.0537 (0.0131)	0.633	0.0395 (0.0074)	0.0317 (0.0146)	0.632
CVD	0.0289 (0.0029)	0.0459 (0.0031)	0.000	0.0212 (0.0031)	0.0362 (0.0041)	0.004
hypertension	0.0104 (0.0025)	0.0141 (0.0028)	0.316	0.0030 (0.0028)	0.0024 (0.0037)	0.908
circulatory problems	0.0128 (0.0039)	0.0166 (0.0044)	0.527	0.0081 (0.0035)	0.0094 (0.0050)	0.840
bronchitis	0.0212 (0.0043)	0.0299 (0.0059)	0.234	0.0153 (0.0041)	0.0121 (0.0068)	0.684
asthma	0.0066 (0.0051)	0.0130 (0.0067)	0.449	−0.0026 (0.0048)	−0.0064 (0.0076)	0.666
emphysema	0.0605 (0.0088)	0.0752 (0.0069)	0.189	0.0415 (0.0081)	0.0590 (0.0080)	0.122
other lung disorder	0.0321 (0.0092)	0.0710 (0.0090)	0.003	0.0171 (0.0084)	0.0502 (0.0102)	0.012
respiratory cancer	0.3469 (0.0264)	0.3683 (0.0194)	0.513	0.3342 (0.0236)	0.3545 (0.0214)	0.524

Notes: The results shown in the first three columns are from linear regressions of an indicator that the respondent died within 2 years of the survey on an indicator that the respondent had the condition listed in the row, plus a set of age indicators, indicators for race (black or other non-white), and years of education. Separate regressions were estimated for each condition. Each cell shows the regression coefficient and standard error for the condition listed in the row. The third column is the p-value for a t-test of the hypothesis that the effects of each condition on male and female mortality are identical. The results in the last three columns are from regressions of an indicator that the respondent dies within 2 years of the survey on a set of indicators for whether the respondent had each of the conditions, plus a set of age and race indicators and a control for years of education. These estimates have been corrected to account for the fact that not all conditions are observed for each individual, using the procedure described in the text and in the Appendix.

Table 5: Reports of chronic conditions for those with specific causes of death

Cause of death	Condition	Women		Men		χ^2 (p-value)
		_N_	% with condition	_N_	% with condition	
Digestive cancer	Digestive cancer	363	3.31%	384	5.21%	1.65 (0.199)
Diabetes	Diabetes	177	51.4%	141	54.6%	0.32 (0.164)
Heart attack	CVD	544	32.4%	669	35.9%	1.65 (0.199)
Stroke	CVD	393	33.8%	304	35.5%	0.21 (0.643)
Hypertensive disease	Hypertension	122	60.7%	93	49.5%	2.68 (0.102)
Respiratory disease	Asthma, bronchitis, or emphysema	267	41.6%	319	54.2%	9.33 (0.002)
Respiratory cancer	Respiratory cancer	15	3.91%	31	5.51%	1.26 (0.261)

Notes: Each row of this table is based on a sample of respondents who died within 2 years of the survey and had the specific cause of death listed in the first column. The table shows the percentages (of those who died from each cause) who reported the condition listed in column 2. The χ^2 test shows whether the percentages of women and men who reported the condition differ. The causes of death are from the NHIS 72-item ICD-9 recode, and are defined as follows: Digestive cancer= Malignant neoplasms of digestive organs and peritoneum. Diabetes= Diabetes mellitus. Heart attack= Acute myocardial infarction.Stroke= Intracerebral and other intracranial hemorrhage; Cerebral thrombosis and unspecified occlusion of cerebral arteries; cerebral embolism; and "all other and late effects of cerebrovascular diseases." Hypertensive disease= Hypertensive heart disease; Hypertensive heart and renal disease; and Hypertension with or without renal disease. Respiratory disease= Bronchitis, chronic and unspecified; Emphysema; Asthma; and "Other chronic obstructive pulmonary diseases and allied conditions." Respiratory cancer=Malignant neoplasms of respiratory and intrathoracic organs.

Table 6: Averages of risk factors for coronary heart disease

	Cardiovascular disease at baseline (by self report)			
	men	women	men minus women	t-statistic, men minus women=0
Indicator: Severe chest pain lasting more than ½ hour	0.396	0.336	0.060	1.68
Indicator: Current smoker	0.321	0.245	0.076	2.27
Number of heart attacks ever	1.094 (1.000)	0.687 (1.033)	0.407	5.43
Indicator: High systolic blood pressure (≥ 120 mm Hg)	0.740	0.776	–0.036	1.12
Indicator: High diastolic blood pressure (≥ 80 mm Hg)	0.398	0.242	0.156	4.51
Indicator: High total cholesterol (≥ 240 mg/dL)	0.263	0.417	–0.155	4.53
Indicator: High triglycerides (≥ 150 mg/dL)	0.574	0.619	–0.046	1.26
Indicator: High risk HDL cholesterol (≤ 40 mg/dL)	0.470	0.243	0.227	6.51
Framingham 10-year-risk of CHD	0.196 (0.119)	0.119 (0.078)	0.077	9.96
observations	440	329		

Notes: The data consist of all men and women ages 30-74 from the NHANES III, who self-reported having cardiovascular disease (have had a stroke, heart attack or congestive heart failure.) Averages are computed using sample weights provided in NHANES III. Standard deviations are in parentheses.
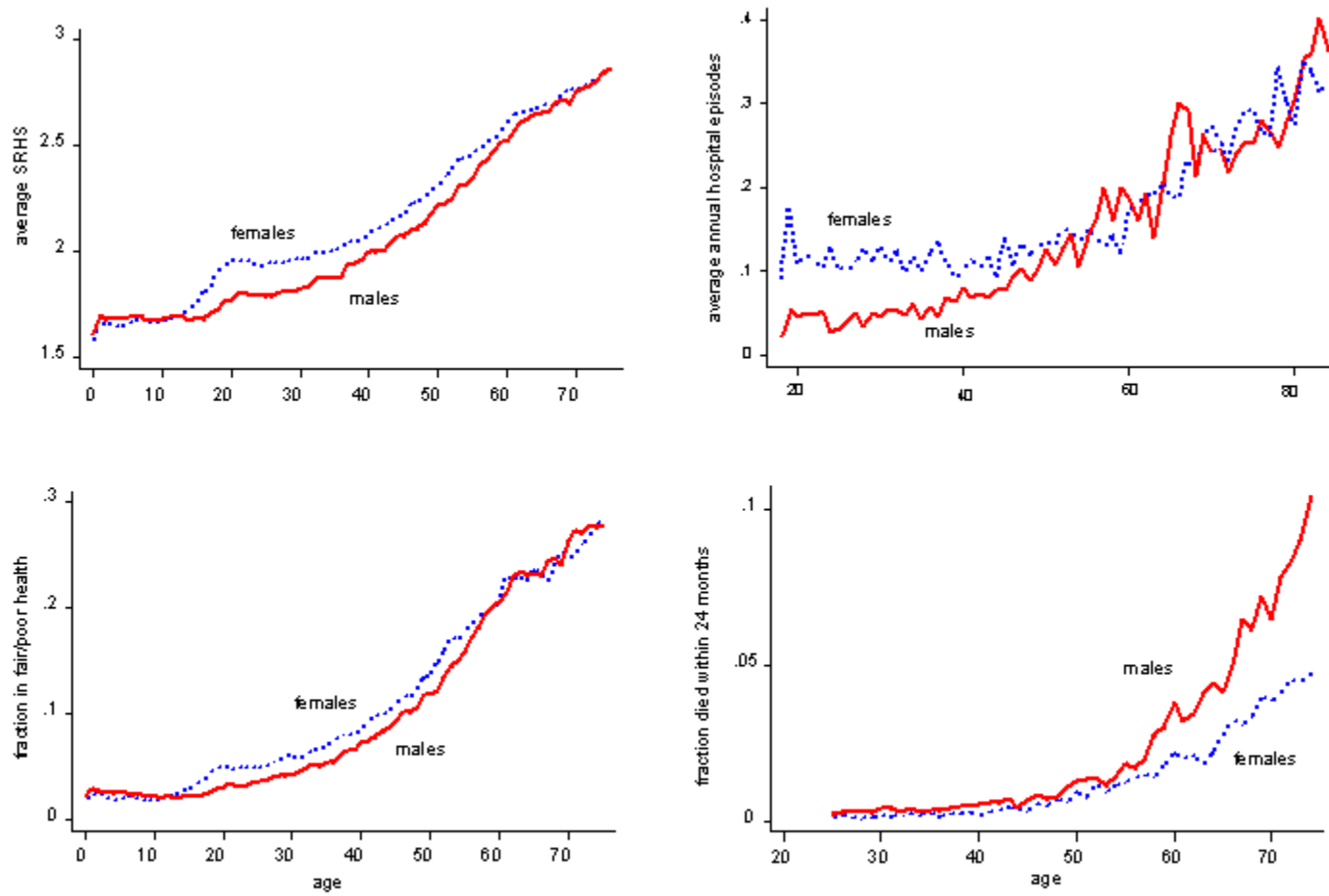
Figure 1: Self-rated health, hospital episodes and mortality for men and women. NHIS.
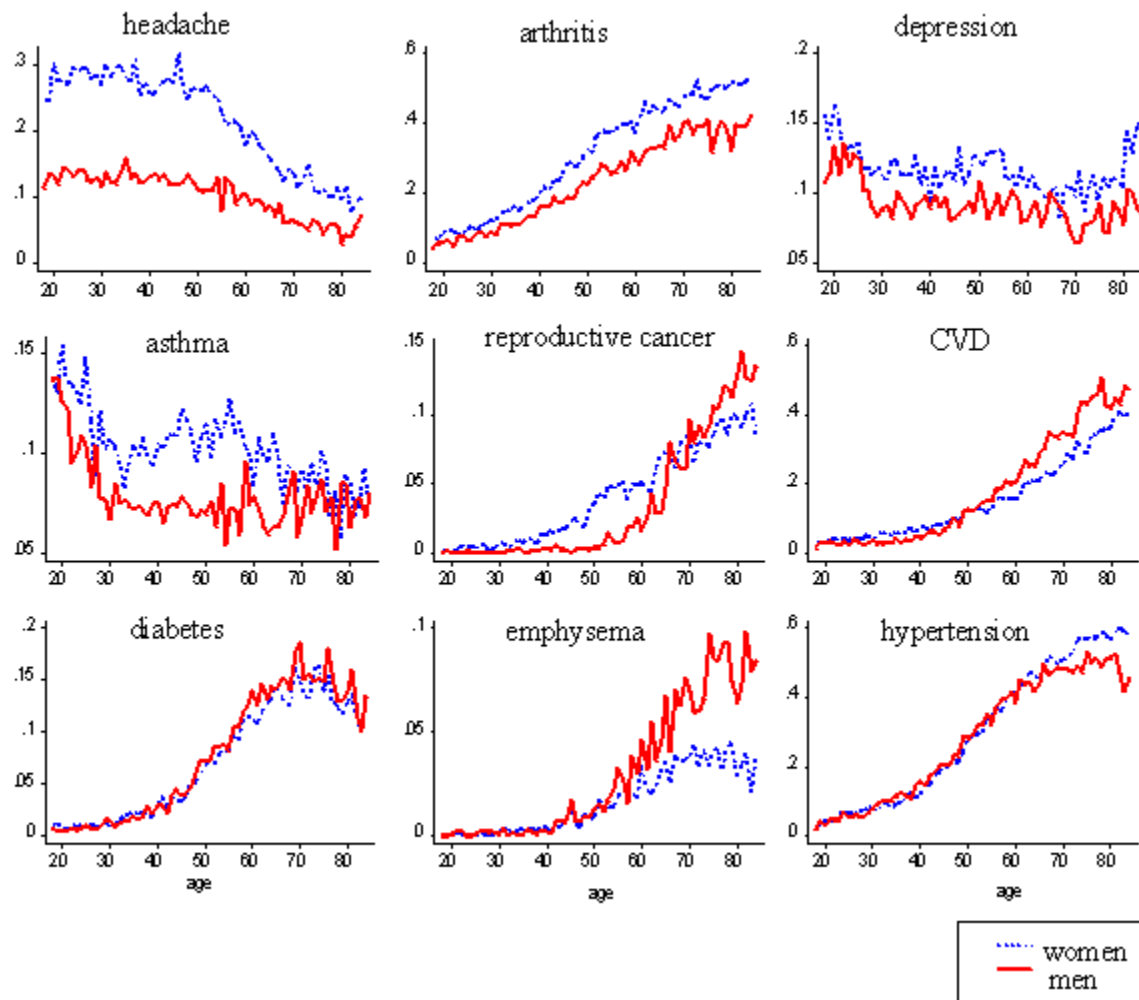
34

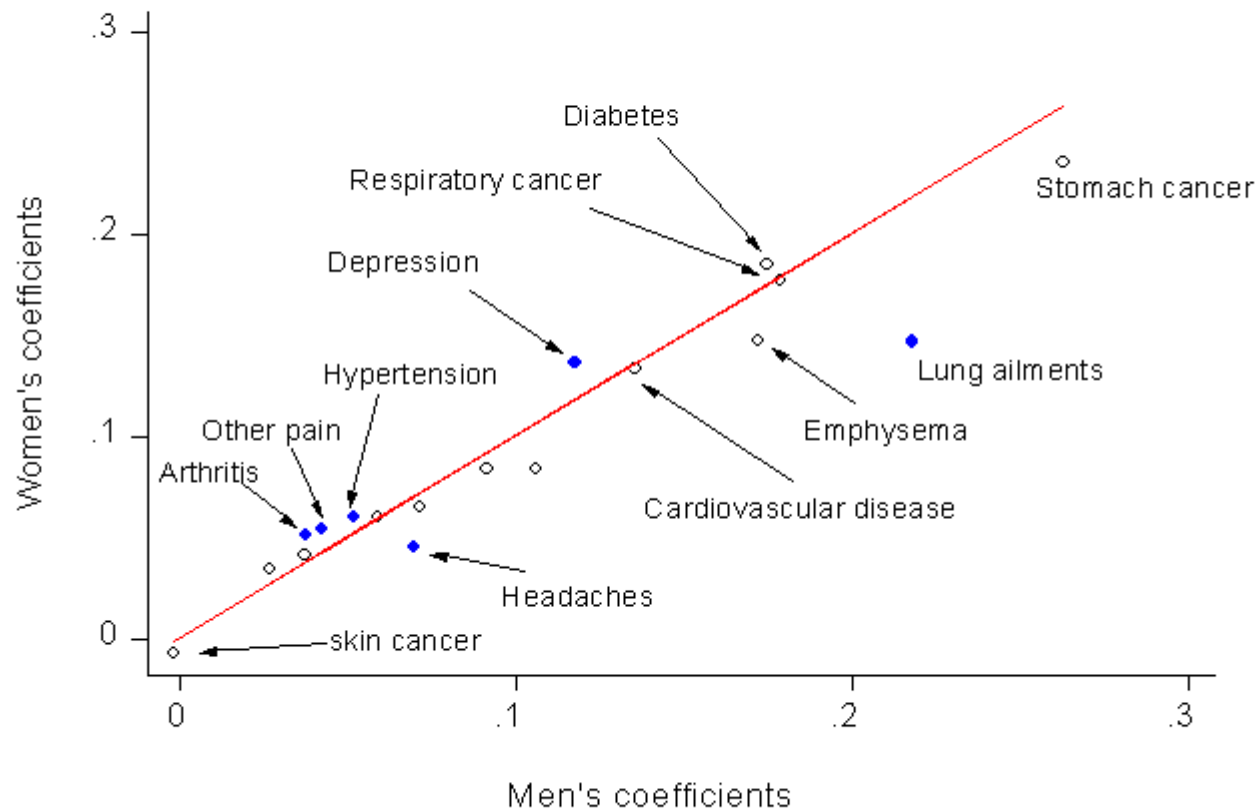**Figure 2**: Prevalence of selected conditions by age

Figure 3: The impact of chronic conditions on the probability of reporting fair or poor health
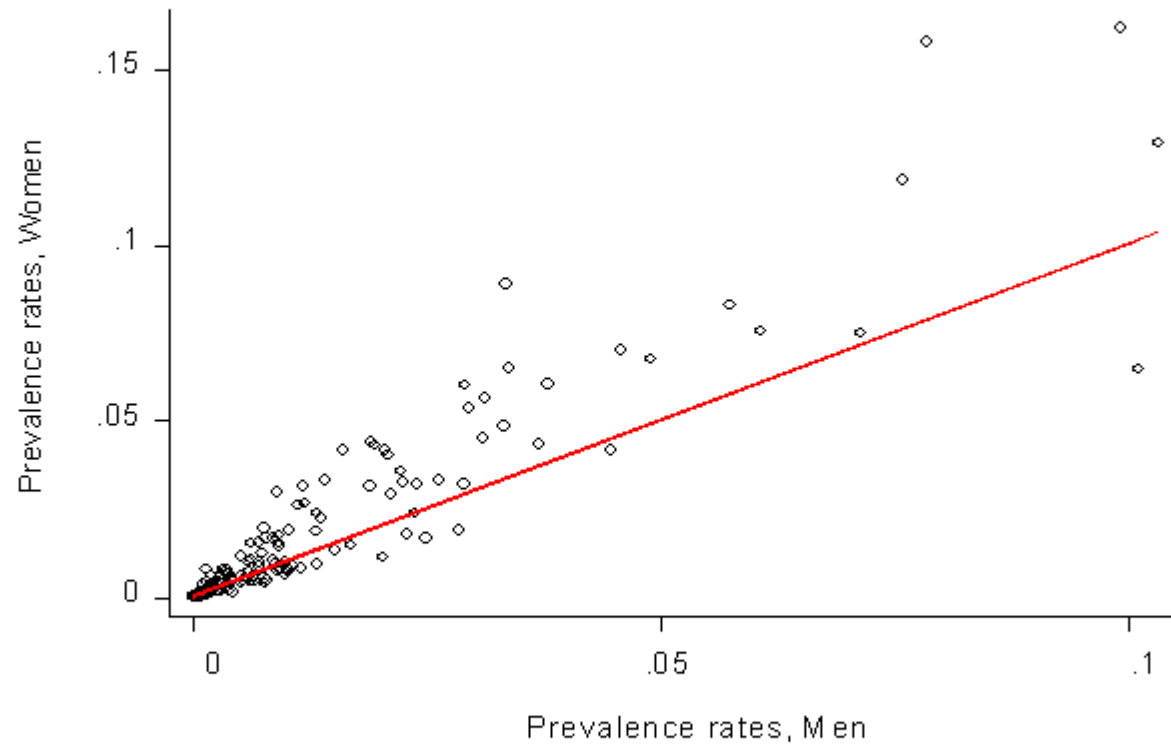(Coefficients significantly different between men and women are represented by solid circles.)

Figure 4: The prevalence of 171 chronic conditions and condition interactions. Ages 18-64. NHIS 1997-2001.
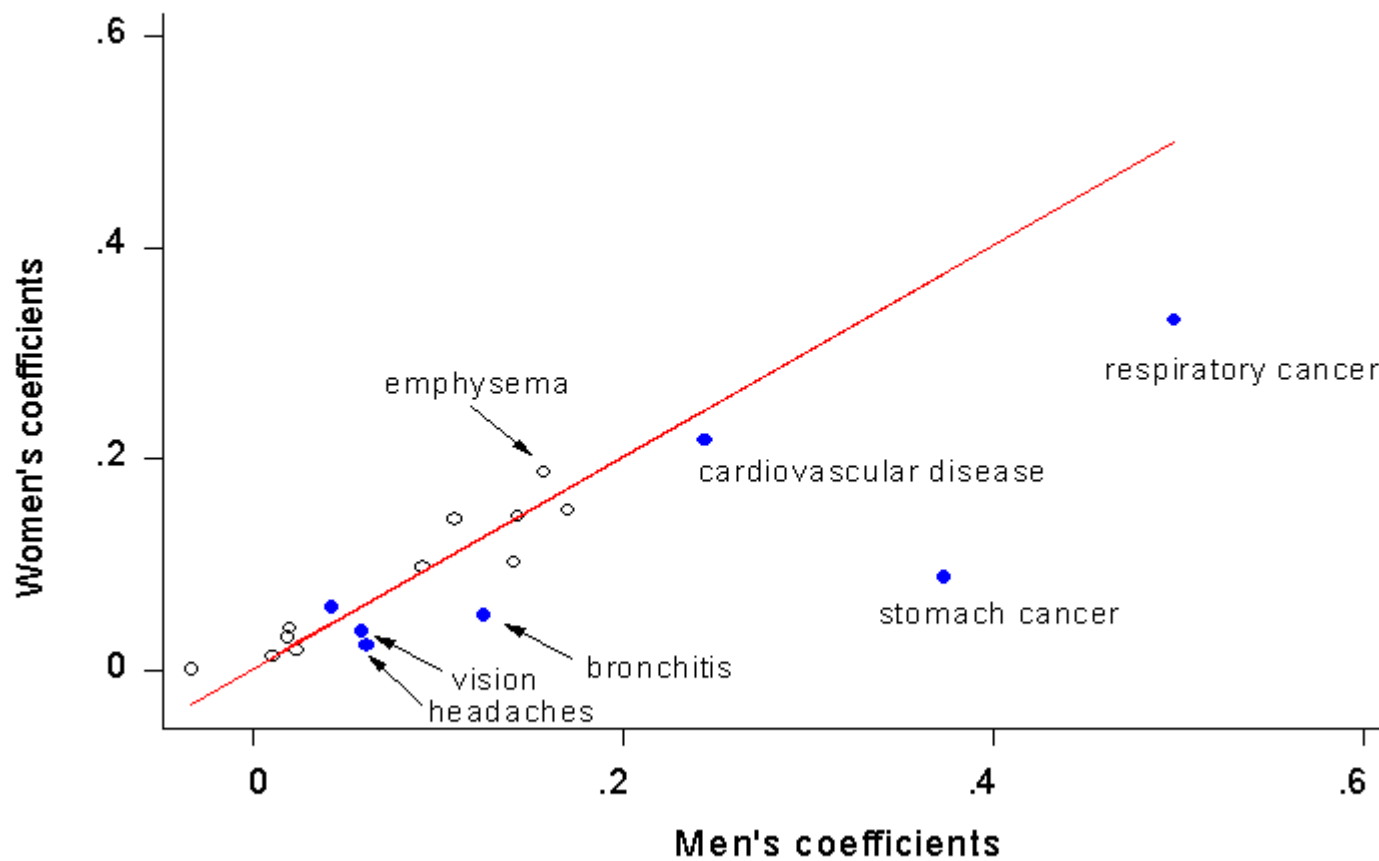
Figure 5: The impact of chronic conditions on hospital episodes in past 12 months.
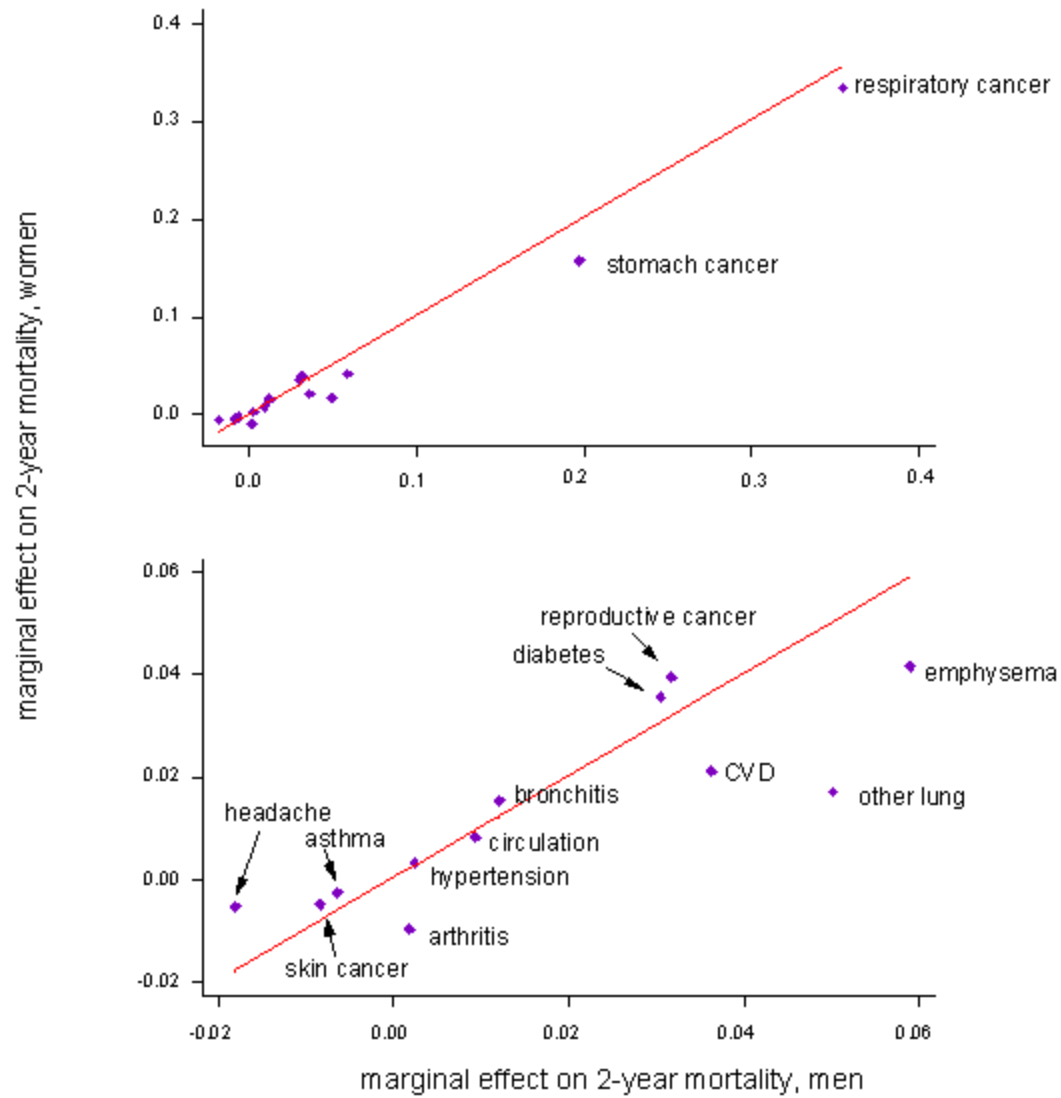(Coefficients significantly different between men and women are shown as sold circles.)

Figure 6: Marginal effects of health conditions on 2-year mortality, men and women.
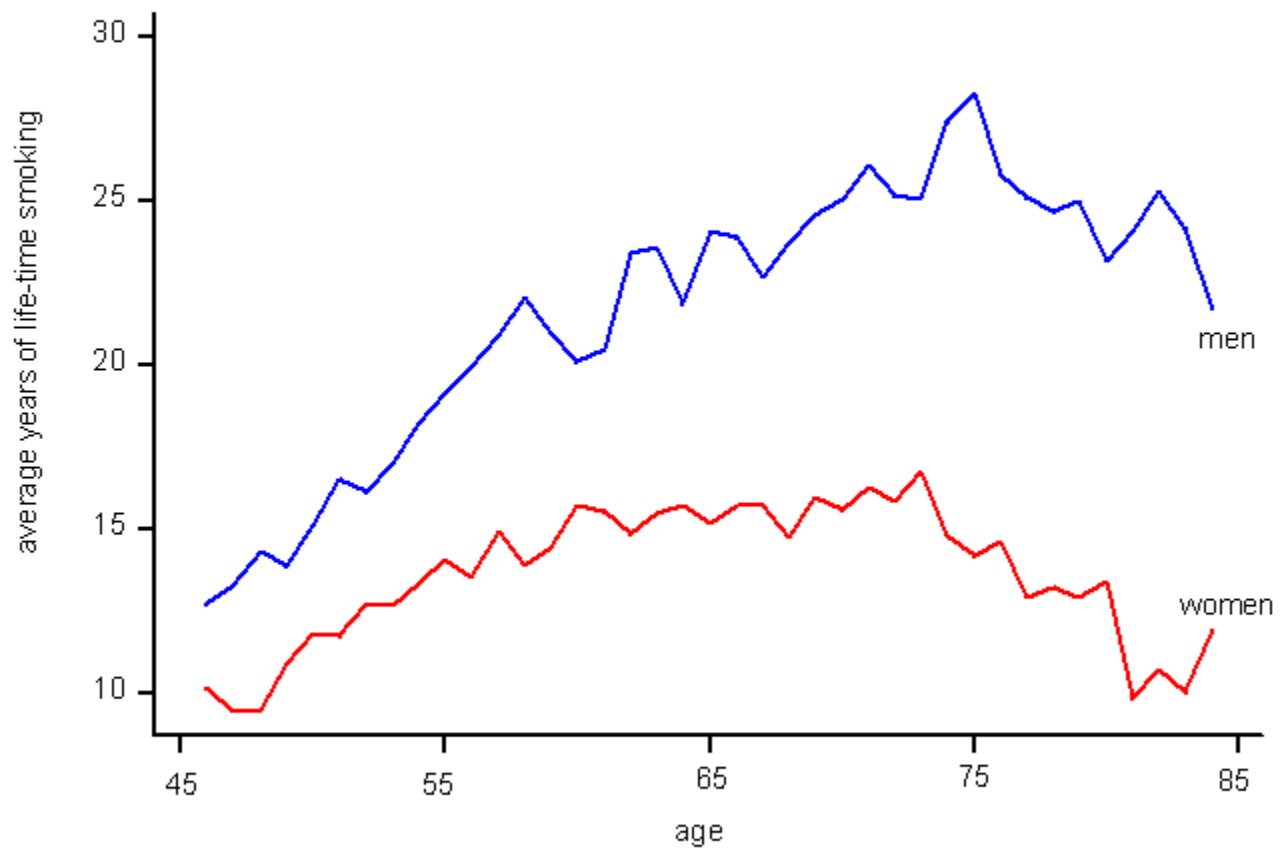
Figure 7: Average years of life-time smoking for men and women, by age. 1997-2001 NHIS.

Appendix

A. <u>Definitions of the variables</u>

All regressions included a complete set of indicators for age and race (black, white or "other") and a control for years of education. The chronic health conditions used in the analyses are defined as follows for each of the two survey periods. In coding the conditions, we attempted to make the definitions for conditions from the earlier and later periods as close as possible. For the 1986-1994 surveys, we provide the diagnostic codes that were aggregated to create each condition. The reference periods over which conditions are assessed are noted for the 1997-2001 period :

1986-1994. 12 month reference period unless noted otherwise.	
arthritis	arthritis (101); rheumatis, unspecified (102); gout, incl. gouty arthritis (103)
skin cancer	malignant neoplasms of the skin (119)
digestive cancer	malignant neoplasms of the stomach, intestines, colon or rectum (316)
reproductive cancer	malignant neoplasms of breast (421), female genital organs (422), or prostate (423)
cardiovascular disease	rheumatic fever "ever" (501), ischemic heart disease "ever" (502); tachycardia or rapid heart (503); heart murmur (504); other unspecified heart rhythm disorder (505); congenital heart disease "ever" (506); other selected diseases of the heard excl hypertension (507); cerebrovascular disease "ever" (509); hardening of the arteries "ever" (510); aneurysm (511)
hypertension	hypertension "ever" (508)
circulation problems	Phlebitis, thrombophelitis (512); hemmorhoids (514); poor circulation (515)
bronchitis	chronic bronchitis (601)
asthma	asthma (602)
emphysema	emphysema (609)
other lung disorders	pleurisy (610); pneumoconiosis and asbestosis (611); tuberculosis (612); other diseases of the lung (614)
respiratory cancer	malignant neoplasms of the lung or bronchus (613); malignant neoplasms of other respiratory sites (615)
1997-2001	
headache	Severe headache/facial pain in last 3 months
neck/backache	Pain in neck, lower back, or joint injury in last 3 months
arthritis	Joint pain not due to injury in last 12 months.
cardiovascular disease	"Ever been told" has had heart attack, coronary heart disease, angina, other kind of heart condition/heart disease, stroke.

diabetes	"Ever been told" has diabetes.
hypertension	"Ever been told" has hypertension.
circulatory problems	Difficulty with activities due to a "circulatory problem"
chronic vision problems	Trouble seeing "even when wearing glasses"
chronic hearing problems	Wears hearing aid, has "a lot of trouble" hearing, or reports being deaf.
depression	Respondent were asked to report whether they felt sad, nervous, restless, hopeless, that everything was an effort, or worthless over the past 30 days. Individuals who reported having at least one of these feelings either "all of the time" or "most of the time" are coded as depressed.
bronchitis	"Past 12 months" has been told has chronic bronchitis.
emphysema	"Ever been told" has emphysema.
asthma	"Ever been told" has asthma
other lung problems	Difficulty with activities due to a "lung or breathing problem"
skin cancer	"Ever diagnosed" with skin cancer.
stomach cancer	"Ever diagnosed" with stomach cancer
reproductive cancer	"Ever diagnosed" with reproductive cancer.
respiratory cancer	"Ever diagnosed" with respiratory cancer.

B. Bias corrections for mortality estimates

The equation to be estimated is:

$$D = C_0 \beta_0 + C_1 \beta_1 + \dots + C_M \beta_M + \epsilon \ = C\beta + \epsilon, \tag{A1}$$

where it is assumed that ϵ is *i.i.d.* with $E[\epsilon \mid C]{=}0$. C_0 is a matrix of demographic variables (age, education and race) that are observed for all individuals. For $j{=}1\dots M$, C_j is an N x k_j matrix of 0/1 variables that indicate whether the individual has each of the conditions included in condition list j, where there are a total of M condition lists. Assume that the number of individuals who are asked about the conditions in each list is equal to $n{=}N/M$, and that the total number of explanatory variables is $\sum_{j=0}^{M} k_j = k$. The matrix C is censored, so that for individuals assigned to condition list j, only variables in C_0 and C_j are observed. The outcome D is observed for all individuals.

Define a matrix Z which is equal to C but with censored elements set to zero, so that:

$$C = \begin{bmatrix} C_{01} & C_{11} & \cdots & C_{M1} \\ C_{02} & C_{12} & \cdots & C_{M2} \\ \cdot & \cdot & \cdots & \cdot \\ C_{0M} & C_{1M} & \cdots & C_{MM} \end{bmatrix} \quad \text{and} \quad Z = \begin{bmatrix} C_{01} & C_{11} & 0 & \cdots & 0 \\ C_{02} & 0 & C_{22} & \cdots & 0 \\ \cdot & \cdot & & \cdots & \cdot \\ C_{0M} & 0 & 0 & \cdots & C_{MM} \end{bmatrix}$$

where C_{ij} is an n x k_j matrix of information for individuals assigned to condition list j.

In what follows, it will be useful to have probability limits of the matrices $(C'C)/N$, $(Z'C)/n$, and $(Z'Z)/n$. Assume that:

$$\underset{N \to \infty}{plim} \; \frac{C'C}{N} = \Sigma = \begin{bmatrix} \Sigma_{00} & \Sigma_{01} & \cdots & \Sigma_{0M} \\ \Sigma_{10} & \Sigma_{11} & \cdots & \Sigma_{1M} \\ \cdot & \cdot & \cdots & \cdot \\ \Sigma_{M0} & \Sigma_{M1} & \cdots & \Sigma_{MM} \end{bmatrix} . \tag{A2}$$

Then:

$$\underset{N \to \infty}{plim} \; \frac{Z'C}{n} = \Sigma_{ZC} = \begin{bmatrix} M\Sigma_{00} & M\Sigma_{01} & \cdots & M\Sigma_{0M} \\ \Sigma_{10} & \Sigma_{11} & \cdots & \Sigma_{1M} \\ \cdot & \cdot & \cdots & \cdot \\ \Sigma_{M0} & \Sigma_{M1} & \cdots & \Sigma_{MM} \end{bmatrix} . \tag{A3}$$

and:

$$\underset{N \to \infty}{plim} \; \frac{Z'Z}{n} = \Sigma_{zz} = \begin{bmatrix} M\Sigma_{00} & \Sigma_{01} & \Sigma_{02} & \cdots & \Sigma_{0M} \\ \Sigma_{10} & \Sigma_{11} & 0 & \cdots & 0 \\ \Sigma_{20} & 0 & \Sigma_{22} & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & 0 \\ \Sigma_{M0} & 0 & 0 & \cdots & \Sigma_{MM} \end{bmatrix} . \tag{A4}$$

Finally, assume that there is an supplemental data set that contains N_S observations on all variables, with the $N_S \times k$ data matrix denoted C_S. Assume that:

$$\plim_{N_S \to \infty} \frac{C_S' C_S}{N} = \Sigma \, .$$

The starting point of the bias correction is the OLS estimator of β using Z in place of C, i.e. $\tilde{\beta} = (Z'Z)^{-1} Z'D$. The bias-corrected estimate of β shown in equation (7) is:

$$\hat{\beta} = \hat{\Sigma}_{ZC}^{-1} \hat{\Sigma}_{ZZ} \tilde{\beta} \, . \tag{A5}$$

The estimate of $\hat{\Sigma}_{ZZ}$ is obtained from the censored data as $(Z'Z)/n$. Note that, as in (A4), the lower right portion of this matrix is block diagonal, with estimates of the matrices Σ_{jj} for j=1..M along the diagonal. Note that the elements of these matrices are equal to:

$$\hat{\Sigma}_{jj}(i,k) = P_{ik}^j$$

where P_{ik}^j is the fraction of individuals assigned to condition list j who have both the ith and kth condition within that list. When i is equal to k, this is simply equal to the prevalence of the condition.

The matrix Σ_{ZC} is estimated using the censored and supplemental samples. Specifically, cross-products that are based on within-condition-list terms are taken from the censored sample, and are identical to the corresponding blocks in the estimate of Σ_{ZZ}. The off-block-diagonal terms that represent cross-products that cross condition lists are estimated as follows:

$$\hat{\Sigma}_{jl}(i,k) = P_{ik} = P_i P_k + \rho_{ik} \sqrt{P_i P_k (1-P_i)(1-P_k)} \tag{A6}$$

where P_i and P_k equal the fraction of the censored sample that has condition i and k, respectively, and ρ_{ik} is the correlation between condition i and k that is estimated from the supplemental sample. An alternative method of obtaining $\hat{\Sigma}_{jl}(i,k)$ would have been to compute the fraction of the supplemental sample that has both condition i and k. However, for some pairs of rare conditions, we found that the fraction of the supplemental sample that had both condition i and condition k exceeded the fraction of the censored sample that had either i or k. The use of (A6), which draws only the correlation coefficient between i and j from the supplemental sample, prevents this from occurring.

The proof that $\hat{\beta}$ is consistent is straightforward. Substituting the formula for $\tilde{\beta}$ into (A5), we get:

$$\hat{\beta} = \hat{\Sigma}_{ZC}^{-1} \frac{Z'C}{n} \beta + \hat{\Sigma}_{ZC}^{-1} \frac{Z'\epsilon}{n}$$

the first term of which has a probability limit of β and the second term of which has a probability limit of 0.

The estimate of the variance-covariance matrix for $\hat{\beta}$ is:

$$\frac{\sigma^2}{n} \hat{\Sigma}_{ZC}^{-1} \hat{\Sigma}_{ZZ} \hat{\Sigma}_{ZC}^{-1} \tag{A7}$$

where σ^2 is the variance of ϵ. A consistent estimate of σ^2, denoted $\hat{\sigma}^2$, is obtained by starting with the estimate of the error variance using the error-ridden estimate $\tilde{\beta}$. Specifically, let:

$$\tilde{\epsilon} = D - Z\tilde{\beta} = [I - Z(Z'Z)^{-1}Z']D = M_Z[C\beta + \epsilon] \, .$$

Then:

$$\frac{\tilde{\epsilon}'\tilde{\epsilon}}{N-k} = \tilde{\sigma}^2 = \frac{\beta'C'M_ZC\beta}{N-k} + \frac{2\epsilon'M_ZC\beta}{N-k} + \frac{\epsilon'M_Z\epsilon}{N-k}$$

and:

$$E[\tilde{\sigma}^2] = \frac{\beta'C'M_ZC\beta}{N-k} + \sigma^2 \, . \tag{A8}$$

The estimate $\hat{\sigma}^2$ is obtained by subtracting the first term on the right-hand-side of (A8) from $\tilde{\sigma}^2$. Substituting in for M_Z, this yields:

$$\hat{\sigma}^2 = \tilde{\sigma}^2 - \frac{N}{N-k} [\beta'\hat{\Sigma}\beta - \frac{1}{M}\beta' \hat{\Sigma}_{ZC} \hat{\Sigma}_{ZZ}^{-1} \hat{\Sigma}_{ZC}\beta]$$

where $\hat{\Sigma}$ is constructed from $\hat{\Sigma}_{ZC}$. Specifically, where $\hat{\Sigma}$ equals $\hat{\Sigma}_{ZC}$ with all elements of the first k_0 rows divided by M (compare A2 and A3, above).