Looking for private information in self-assessed health

James Banks
Thomas Crossley
Simo Goshev

July 2007

Looking for Private Information in Self-Assessed Health

James Banks,
University College London, IFS and SEDAP

Thomas Crossley,
University of Cambridge, McMaster University, IFS and SEDAP

And

Simo Goshev,
McMaster University and SEDAP[*],

July 28, 2007

Abstract:

The paper investigates whether self-assessed health status (SAH) contains information about *future* mortality and morbidity, beyond the information that is contained in standard "observable" characteristics of individuals (including pre-existing diagnosed medical conditions). Using a ten-year span of the Canadian National Population Health Survey, we find that SAH does contain private information for future mortality and morbidity. Moreover, we find some evidence that the extra information in SAH is greater at older ages.

Many developed countries are experiencing a major shift from defined benefit (DB) to defined contribution (DC) pension arrangements. One consequence of this shift is an effective delay in the age at which workers commit to an annuity. Our results therefore suggest that adverse selection problems in annuity markets could be more severe at older ages, and therefore, that the DB to DC shift may expose workers to greater longevity risk. This is an aspect of the DB to DC shift that has received little attention.

Key words: Self-Assessed Health, Annuities, Mortality, Morbidity
JEL Codes: *H0, I1*

---

1. Introduction

The goal of this paper is to investigate whether self-assessed health status (SAH) contains information about *future* mortality and morbidity, beyond the information that is contained in standard "observable" characteristics of individuals (including demographics, risk behaviors, and pre-existing diagnosed medical conditions). To the extent that SAH does have predictive power for future health shocks, we are particularly interested in how that predictive power varies with age. That is, we hope to understand how individual's uncertainty about their future health status resolves as they age, and in particular, whether people have "private information" about their future health status and whether the amount of private information changes with age.

There are a number of reasons to be interested in this question. The information content of SAH, which is easily collected and included in many surveys, is obviously a relevant issue for the great body of empirical work that uses SAH as either an explanatory variable or an outcome measure.

However, one particular reason to be interested in this question is because of the current trend away from defined benefit pensions and towards defined contribution pensions. Much has been made of the fact that this trend exposes workers to greater financial market risk. However, it may also alter worker's exposure to longevity risk, and this aspect of changing pension arrangements has received little, if any attention.

Longevity risk is simply the risk that an individual may live longer than they expect. While this is, of course, a positive surprise, it can pose severe financial difficulties if the individual does not have adequate financial resources for this extra period of life. The obvious way to avoid such difficulties is to annuitize wealth. One way to think about the switch from DC to DB pensions is that individuals in DC pensions annuitize their pension wealth at retirement. In

contrast, individuals in DB pensions effectively lock into an annuity when they join the firm - typically when they are quite young.

It is well know that take-up of private annuities is surprisingly low. There are a number of reasons why this might be the case. One reason could be that annuity markets suffer from significant adverse selection. It could be that only individuals who have private knowledge that their health is good wish to purchase annuities - so the adverse selection problem is the reverse of what one has in health or life insurance.

If individuals have substantially more private information about their health/expected longevity at age 65 than at age 35, the market for annuities at age 65 will suffer from more adverse selection than the market for annuities that are locked in at age 35. Thus the DB to DC switch may make it more difficult for individuals to insure longevity risk. Brugiavini (1993) develops some of these ideas in a formal theoretical model. However, as noted above, this is an aspect of the trend to DC pensions that has not received much attention. This concern of course, rests on the presumption that individuals have more private information about their health at older ages. It is this hypothesis that we examine in this paper.

Our analysis employs a ten-year span of the Canadian National Population Health Survey (NPHS). This unusual panel survey collects detailed health information from respondents every two years, and the initial sample contained a full range of ages (as opposed, for example, to the retirement and aging surveys underway in several countries, which respondents typically only enter after the age of 50.) To preview our results, we find that SAH does contain private information for future mortality and morbidity. Moreover, we find some evidence that the extra information in SAH is greater at older ages.

The next section reviews the relevant literature. Section 3 provides details on the data and

the estimation approach utilized. The results are presented and discussed in Section 4. Section 5 provides a concluding discussion.

2. Literature Review

The introduction of mandatory retirement savings plans and the transition from DB to DC pensions in many developed countries has led to a rapid growth in the private annuity markets in those states. Despite the growth however, those markets have continued to be "not well developed even in the most advanced OECD countries" (James and Vittas 1999). One reason for this observed underdevelopment may be the presence of adverse selection in these markets, and this possibility has been the focus of much recent research.

One approach to the study of annuity markets is to evaluate the "value per premium dollar" of annuities offered for sale (see for example Mitchell et al., 1997). Such studies typically find values significantly below one. The insurance load in excess of reasonable administrative costs is attributed to adverse selection.

An alternative approach to test for adverse selection is to look for correlation between annuity purchases and subsequent realized risk experience. Finkelstein and Poterba (2002) observe that in the UK annuities markets annuitants, particularly voluntary annuitants, live longer than non-annuitants. Moreover, they find that "the pricing of different types of annuity products within each annuity market is consistent with individuals selecting products based, in part, on private information about their mortality prospects". Finkelstein and Poterba (2004) document further evidence of a systematic relationship between future mortality and annuity characteristics.

Finally, Finkelstein and Poterba (2006) construct a test for adverse selection in insurance markets that is potentially able to distinguish adverse selection from moral hazard. The test, based on observable characteristics of insurance buyers that are not used in setting insurance prices, provides evidence of the presence of adverse selection.

The only evidence on adverse selection in Canadian annuity markets that we are aware of is Milevsky (1998). Following the methodology of Mitchell et al (1997), Milevsky calculates value per premium dollar for Canadian annuity quotes in the period 1984-1996. He focuses exclusively on 65-year old men and women and ignores the value available at other ages. Milevsky (1998) finds value per premium dollar of about 90 cents (or, equivalently, an insurance load of about 10%). The estimates vary with alternative assumptions about mortality and the term structure of interest rates. Value per dollar of premium is higher when using annuitant life tables than when using population life tables. This reflects the greater longevity of annuitants implicit in the life tables and is consistent with adverse selection.

All of these studies take the approach of inferring adverse selection from prices or quantities in annuity markets. In this paper, we follow the alternative, and complementary strategy of trying to determine directly whether individuals actually have private information about health and longevity. One reason to take this alternative approach is that it may shed light on whether adverse selection in annuity markets is "active" or "passive". Poterba (2001) mortality differences between annuitants and non-annuitants might arise if there were correlations between the characteristics annuity purchasers and longevity. Moreover, annuitant purchasers need not be aware of these correlations. For example, annuitants tend to be wealthy and have incomes; these factors are plausibly correlated both with annuity demand and with

health and longevity. Thus while differences in the longevity of annuitants establishes that there is selection into annuitant status, it does not establish that this selection arises because of individuals acting on private information. Our approach is to look directly for private information.

The most natural way to do this would be to examine individual's responses to survey questions about their longevity expectations. Smith et al (2000) utilize the U.S. Health and Retirement Survey (HRS) and find that longevity expectations predict mortality at the individual level. Their results also suggest that health shocks and certain health conditions negatively impact longevity expectations. Similarly using the HRS, Hurd and McGarry (2002) look at the evolution of subjective survival probabilities and their ability to predict actual mortality. They find that subjective survival probabilities do predict actual survival.

The problem with studying longevity expectations in the context of our work is that life-expectancy questions have, to date, mostly been asked in retirement surveys. These surveys only collect data from people over the age of 50. Thus these data cannot be used to compare the private information held by younger and older individuals, which is the comparison that we are most interested in.

A potential proxy measure of longevity expectations is self-assessed health (SAH). This measure is widely available and frequently employed in the economics and epidemiology literature on mortality. Therefore, to assess the amount of private information that individuals have, we look at the effect of SAH on future mortality and morbidity while controlling for a rich set of observables including pre-diagnosed health conditions and risk behaviours. The idea is to explore whether SAH contains information beyond that which would typically be available to an annuity seller.

The literature on the predictive power of SAH for future mortality and morbidity is extensive and has established that SAH is a significant predictor of future health outcomes. Early studies (Mossey and Shapiro 1982, Okun et al 1984, McCallum 1994, Idler and Kasl 1995) find that self-rated health predicts morbidity and survival. Idler and Benyamini (1997) summarize results from U.S. and international longitudinal studies on self-assessed health as a mortality predictor. They conclude that despite the differences in methodology and controls, self-assessed health is a recognized globally as an independent predictor of mortality. Schwarze et al (2000) confirm this finding with German data. Several recent studies looking at self-rated health, health care utilization (DeSalvo et al 2005) and hospital episodes (Case and Paxon 2005) find that self-assessed health is a predictor of mortality and that its effect varies by gender and baseline chronic conditions.

To evaluate whether individuals possess more private information about their health at older ages, we need to look at data collected from respondents spanning the entire age range. We then have to estimate the effects of SAH on future mortality, conditional on observables and compare the information contained in the self-reported health measure across ages. Two studies: Burstrom and Fredlund (2001), and Van Doorslaer and Gerdtham (2003) using Swedish data, take a similar approach.

Burstrom and Fredlund (2001) use the annual cross-sectional Swedish Survey of Living Conditions (SSLC) for the period from 1975 to 1997, linked to Sweden's National Causes of Death Statistics (NCDS). They focus on the mortality ratios of death during the follow-up period in relation to self-reported health at the time of interview. The authors utilize a Cox proportional hazards model and find that the mortality rate ratios for persons reporting bad health compared to individuals reporting good health are high at younger ages, but that the effect declines with age.

The second study, Van Doorslaer and Gerdtham (2003), also employs pooled data from the annual SSCL for 1980 through 1986, once again linked to the NCDS. Using a similar Cox proportional hazards framework, Van Doorslaer and Gerdtham also find that "the effect of SAH on mortality risk declines with age".

Both these papers suggest then, that private information about future health outcomes *declines* with age. Nevertheless, these studies are based on a common Swedish data set, and it seems important to revisit this issue with other data. We do so with data from the Canadian National Population Health Survey.

3. Data and Methods

3.1. Survey Details and Sample of Analysis

The Canadian National Population Health Survey, administered by Statistics Canada, is a longitudinal health survey of the Canadian population. The three target populations of the NPHS are household residents in all Canadian provinces[1], residents foreseen to remain longer than six months in health care institutions, and the residents of Yukon and the Northwest Territories[2].

In all provinces except Quebec, the NPHS household component utilizes a stratified two-stage sampling design based almost entirely on the Canadian Labour Force Survey sampling design. In Quebec, the NPHS employs the design of the 1992-93 Enquête sociale et de santé. The final NPHS household sample is created by selecting households from within cluster-dwelling break-outs and then choosing a household member, 12 years old or older, as the longitudinal respondent to be followed over cycles. The survey is biennial and ongoing. The first cycle gathered data for 1994-95. The most recently released cycle, cycle five, contains data for 2002-

---

[1] Excluding populations on Indian Reserves, Canadian Forces Bases and remote areas in Quebec and Ontario.
[2] Excluding populations on Indian Reserves, Canadian Forces Bases and remote areas.

03.

In this study we utilize the health file of the household component of NPHS. The health file contains demographic, socio-economic and comprehensive health-related information about the longitudinal respondent. Interviewing is conducted in-person and by telephone. The percentage of each method varies across cycles and provinces (Statistics Canada, 1996).

There are 17,276 respondents in cycle 1 falling to 14,532 in Cycle 3 and 12,546 in Cycle 5. Total attrition between Cycle 1 to Cycle 5 is 27.4%. The most common reason for attrition is refusal to provide information and it amounts to 61% of all attrition. In addition, however, by Cycle 5, 1279 cycle 1 respondents are deceased. These individuals can potentially be included in our analysis when mortality is the outcome of interest. Item non-response in Cycle 5 varies from 0% to 5%.

As described in greater detail below, our empirical strategy is to model mortality between Cycles 1 and 5, and morbidity at Cycles 3 and 5, as functions of Cycle 1 information (including self-assessed health). When we model mortality our analysis sample comprises 9004 respondents (4516 male and 4488 female) aged 20 to 64 in Cycle 1. Of these 340 are deceased by Cycle 5. The differences between the numbers above (12,546 Cycle 5 respondent and 1279 deceased) and our working sample are due to the initial age restriction and item non-response in Cycle 1. When modelling morbidity, the deceased represent attrition and our sample is restricted further by item non-response in Cycle 5, which varies between 0% and 5% across items. Thus when looking at morbidity, we utilize a sample of 7439 respondents (3326 males and 4113 females).

Throughout we analyze males and females separately. This is consistent with the fact that males and females are treated differently with respect to annuity characteristics and prices in annuity markets

We have conducted standard tests for non-random attrition; these are described below.

3.2. Variables of Interest

Our focus is on the variable self-assessed health. It has five categories: "excellent", "very good", "good", "fair" and "poor" corresponding to the answers to the question: "In general, how would you describe your health?" Table 1 presents the distribution of SAH by gender-age groups. The rates of excellent/very good health reporting steadily decrease with age for both genders. On the other hand, the rates of reporting fair/poor health exhibit a generally increasing pattern.

We consider indicators of mortality and morbidity as health outcome variables. Our analysis of mortality employs a variable that flags all deceased individuals in the period between Cycles 1 and 5. Deaths in the NPHS are confirmed against the Canadian Vital Statistic Database.

While mortality is the relevant outcome for annuities, at younger ages mortality rates are extremely low. Thus we extend our focus to indicators of morbidity. The idea is to look at aspects of morbidity that are strongly associated with mortality. Therefore, we concentrate on conditions that potentially increase the probability of death. The aspects of morbidity we target are the presence of a "major" condition, a "medium" condition, or an "activity restriction".

An individual is identified as having a major condition if s/he is a subject to heart disease, cancer, and/or stroke. This definition is similar to that employed by Smith (1999). An individual is identified as having a medium condition if s/he has diabetes and/or hypertension. These are significant risk factors for major conditions. Activity restriction flags all respondents who because of a physical or mental condition or a health problem are limited (handicapped and/or long-term limited -- limited in the past 6 months) in the kind or amount of activity they can

perform at home, school, work or other. The definitions of all indicators and their prevalence rates are provided in Tables 2 and 3.

All morbidity flags are constructed in terms of current (Cycle 3 or Cycle 5), prevalence. Since we control for Cycle 1 prevalence, we are effectively looking for changes in prevalence between Cycle 1 and Cycles 5 or 3. The questions on which these morbidity flags are based all have the following general format: "Do you have [condition] diagnosed by a health professional?"

Note that current prevalence at Cycle 5 is necessarily less than total prevalence over the entire 10-year period between Cycles 1 and 5 (and similarly for Cycle 3). The discrepancy varies by condition (see Table 3). However, we have repeated all of the analysis described below with morbidity defined as total prevalence over the relevant period, and the results were very similar to those described below.[3]

The set of Cycle 1 controls we employ includes flags for pre-existing health conditions including minor conditions (defined as any health condition but major or medium) in addition to major and medium conditions and activity restrictions. It also includes risk factors (body mass index and indicators of smoking and drinking) as well as a number of socio-economic and geographic characteristics including age, gender, household income, education, marital status, labour force status, mother tongue, region of residence in Canada. Summary statistics for socioeconomic control variables are provided in Table 4.

3.3. Estimation Strategy and Methodology

Our estimation strategy is as follows. First, we divide the data into age groups: 20-34, 35-49, and 50-64. Then, within each group, we estimate econometric models of the form:

---

[3] Full results are available from the authors.

$$prob(y^j_{t+k} = 1) = f_A(SAH_t, Z_t, y^1_t, ..., y^j_t, ... y^J_t)$$

where y^j_t is a measure of mortality or morbidity at time t; SAH_t is self-assessed health status at time t; and Z_t is a set of observable characteristics. These last would include demographics (age and sex, marital status); socioeconomic variables (education, occupation, income groups) and risk behaviours (smoker or not).

Thus, again, we are testing whether SAH has additional predictive power for future mortality and morbidity once we control for the types of information that would typically be observable by a seller in an annuity or insurance market: demographics, socioeconomic status, some risk behaviours and previously diagnosed conditions ($y^1_t ... y^j_t ... y^J_t$). To determine whether private information about health accumulates with age, we compare estimates of the effect of SAH in models of this type estimated for different age groups (as indicated by the *A* (age) subscript on the function *f*).

The particular functional form we use for *f* is a logit model. From the parameter estimates, we construct two measures of the magnitude of any effect of SAH on the probability future health outcomes. The first is the *marginal effect*. This is the difference between the probability of a future health event for individuals in one SAH category and the probability of the same health event for individuals in another SAH category, measured in *percentage points*. Thus it is an *absolute risk effect*. The second is the odds-ratio minus unity: unity subtracted from the ratio between the probability of a future health event for individuals in one SAH category and the probability of the same health event for individuals in another SAH category. Roughly, this measures the difference in risk across the two groups as a *percentage* of the risk of the base group. Thus it is a *relative risk effect.* The absolute and relative effects are reported separately below. Note that, across age groups, the absolute and relative effects can move in opposite

directions. For example, the absolute effect could increase with age, while the relative effect falls. This would happen if the baseline risk rose faster with age than the absolute effect.

4. Results

We first ask whether SAH has incremental predictive power for mortality. We focus initially on the ten-year time horizon spanned by Cycles 1 and 5. Marginal effects are presented in Table 6 for males and Table 7 for females. Marginal effects of very good or excellent SAH versus a baseline of good health are given in the first row of each table. Marginal effects of fair or poor health, again versus the baseline middle category of good health, are given in the second row. The results for the pooled sample (ages 20 to 64) are given in the first column. Table 6 indicates that male respondents reporting excellent/very good health in Cycle 1 are 1.5 percentage points less likely to experience death over the next 10 years, compared to males reporting good health and controlling for pre-existing conditions, risk factors, and socioeconomic variables. The corresponding odds-ratio, reported in Table 8, indicates that males who report excellent or very good health are approximately one third less likely to experience death over the following 10-year period (as indicated by an odds ratio of 0.66). Both absolute and relative effects are statistically significant at conventional levels ($p<0.05$). Men who report fair or poor health are more likely to die over the subsequent 10 years (again relative to the base group reporting good health, and controlling for initial conditions, risk factors and socioeconomic characteristics) but the effect is not statistically significant (whether measured absolutely or relative to the baseline risk).

Table 9 indicates that women that reported fair or poor health are 65% more likely to experience death, and this effect is statistically significant at the $p < 0.1$ level. However, the

corresponding marginal (or absolute risk) effect (reported in Table 7) is not statistically significant, nor is either the absolute risk or relative risk effect of reporting very good or excellent health.

We next estimate our predictive models separately for the 20-34, 35-49 and 50-64 age groups to investigate whether the incremental predictive power of SAH varies with age. In each of Tables 6, 7, 8 and 9, results for the 20-34 age group are in the second column; results for the 35-49 age group are in the third column; and results for the 50-64 age group are in the fourth and final column. Comparisons of marginal effects for each age group are made graphically in Figures 1 and 2 (for men) and Figures 3 and 4 (for women).

For men, the marginal effect on mortality risk of reporting excellent or very good health (Table 6) is actually positive (though not statistically different from zero) for the youngest group, turns negative (but again not statistically different from zero) for the middle group and is negative and statistically different (at $p<0.01$) for the oldest group. Thus the effect noted in the pooled sample appears to be driven largely by the oldest group. Table 6a reports tests of equality between marginal effects in different groups, and confirms that the marginal effect for the oldest group of men is statistical different from the estimated effect for the youngest ($p = 0.003$) and middle ($p = 0.021$) groups. The marginal effect of poor or fair health is marginally significant in the middle group, but not elsewhere (Table 6) and the effects for different age groups are not statistically different from each other (Table 6a).

When we present the effects in relative (odds ratio) form, in Table 8, the same finding is apparent for very good or excellent: the predictive power observed in the full sample appears to be largely driven by the oldest group. For this group, but not for the younger groups, the odds ratio is strongly statistically different from one. The effects of poor or fair health present a less

interpretable pattern (as they did in when presented as absolute marginal effects). The strongest effect here is for those aged 35 to 49.

The age-group results for the female sample are in the second through fourth columns of Tables 7 (marginal effects) and 9 (odds-ratios). Corresponding tests of equality of marginal effects across age groups are presented in Table 8. None of the within group-age effects (either relative risk or absolute risk) are statistically significant, at even the $p < 0.1$ level. In part this may reflect that the baseline mortality risk is very low, and about half of male risk in these age groups (see Table 5). This means that we are modelling a rare event.

We next ask whether SAH predicts future morbidity, and particularly the emergence of conditions that are associated with mortality risk. The results follow the same pattern as for mortality. Results for males are presented in Tables 6, 6a and 8; for females in Tables 7, 7a and 9. Marginal effects, capturing a difference in absolute risk, are presented in Tables 6 and 7, and Tables 6a and 7a report tests of the equality of marginal effects across age groups. Odds-Ratios, which capture differences in relative risk, are reported in Tables 8 and 9. Moving down each table from the mortality results, we present in turn results for major conditions (heart disease, cancer and stroke), medium conditions (diabetes and hypertension) and activity restrictions.

Beginning with the male sample, and marginal effects, we see that the effect of excellent or good health on morbidity is negative, as expected, and there is some evidence that the magnitude of these effects increases with age. The effect in the pooled (20-64) sample is statistically significant at $p < 0.01$ for medium conditions and activity restrictions, but not for major conditions.

One reason that the pooled estimate for major conditions is not statistically different from zero is that it is *positive* and statistically significant for the youngest (20-34) group. This result

says that, controlling for pre-existing conditions and risk factors, a young man who reported that he was in very good or excellent health was more likely to have a major condition ten years later than a young man that reported good health. This is a surprising result, although the corresponding effect on mortality, discussed above, has the same sign (though is not statistically different from zero). A young man who reported that his health was fair or poor was also statistically more likely to develop a major condition so there is no simple gradient here. At older ages reporting very good or excellent is associated with lower future incidence of a major condition, though the effect is never statistically significant.

For medium conditions and activity restrictions, the point estimate of the effects of reporting very good or excellent health are larger (that is, more negative) in the older age groups. However, though they are not always statistically different from zero, and, as Table 6a illustrates, the precision with which age-group-specific effects are estimated is not sufficient to allow them to be formally distinguished from each other.

As with mortality, the effects of reporting fair or poor health are less clear – very few of the estimated effects are statistically different from zero.

Turning to women, reporting very good or excellent health has a negative and statistically significant effect on the probability of having a major condition or activity restriction 10 years later. In both cases, when broken down by age, the largest and only statistically significant effect is observed in the oldest (50-64) age group. For activity restrictions and medium conditions, reporting a fair or poor health has a statistically significant effect.

The odds-ratios, or effects on the relative risk, presented in Tables 8 (for men) and 9 (for women), tell a similar story. Some of the odds-ratios are extremely large, which reflects the very low baseline risk of some conditions in some age-groups (for example, major conditions among

20-34 year-olds).

We would summarize these results as follows. First, for both men and women, SAH predicts future mortality and morbidity. Second, on balance the predictive power is stronger at older ages. This is true whether we look at absolute or relative risks (which is important because the baseline risks increase with age.)

We repeated the analysis just described but using a six-year (Cycle 1 to Cycle 3) rather than ten-year time horizon. We did this for two reasons. First, it provides a general check on the robustness of our results and some sense of the time scale over which the predictive power of SAH is operative. The six-year and ten-year horizon results are compared graphically in Figures 5 and 6. A summary would be that the six-year horizon results exhibit similar patterns to the ten-year horizon results but are generally weaker. The second reason to move to a six-year horizon is that it allows us to employ the subsequent cycles to do some testing for effects of non-random attrition, following the suggestion of Verbeek and Nijman (1992). Specifically, we augment the six-year models with dummy variables capturing future attrition (attrition between Cycles 3 and 5). The results do not contain any evidence that attrition is a serious problem in our analysis. The attrition dummies are very occasionally significant and if anything, our main results appear to strengthen with their inclusion.[4]

5. Discussion

In this paper we investigate whether self-assessed health status contains information about future mortality and morbidity, beyond the information that is contained in commonly observable characteristics of individuals. Using a ten-year span of the Canadian National Population Health Survey, we find that even after controlling for pre-existing conditions,

---

[4] Full results are available from authors on request.

socioeconomic characteristics, and a range of risk factors, self-assessed health predicts future mortality and morbidity. Moreover, we find some evidence that this effect strengthens with age. We interpret these findings as supportive of the idea that individuals have private information about their likely future health and lifespan. This in turn suggests that the apparent adverse selection in annuity markets could be at least in part "active". Individuals do seem to be aware of private information that might inform their demand for annuity products. Moreover, we find some evidence that the predictive power of SAH strengthens with age. As Brugiavini (1993) has suggested, this means that any change in pension arrangements that effectively delays the commitment to annuitize may carry with it the cost of exacerbated adverse selection.

There are a number of important ways that this research could be extended. First, our reading of the age patterns in the predictive power of SAH in Canadian data differs from results obtained by Burstrom and Fredlund (2001) and Van Doorslaer and Gerdtham (2003) with Swedish data. It is difficult to determine whether the contrast reflects a true difference in the underlying populations, or differences in the way SAH is measures across the two surveys, or some other aspect of the data and modelling. Further results from additional data sets would help to resolve the generality of these findings.

Second the NPHS could be further exploited to look at the co-evolution of SAH and diagnosed conditions through life. In particular, we are interested in understanding what events trigger revisions of SAH.

Finally, we have reported the surprising finding that at young ages, excellent/very good SAH, conditional on observables, leads to an increased risk of mortality/morbidity in the male sample. If this result is robust, it might reflect misperceptions leading to underinvestment in health or greater engagement in risky activities. This also warrants further investigation.

References

Adams, P., Hurd, M., McFadden, D., Merrill, A., and Ribeiro, T., (2003), Healthy, Wealthy, and Wise? Tests for Direct Causal Paths between Health and Socioeconomic Status, *Journal of Econometrics*, 112(1), pp. 3-56

Brugiavini, A., (1993), Uncertainty Resolution and the Timing of Annuity Purchases, *Journal of Public Economics*, vol. 50, pp. 31-62

Deaton, A., (1999), Inequalities in Income and Inequalities in Health, NBER Working Paper No. 7141

Banks, J., and Blundell, R., (2005), Private Pension Arrangements and Retirement in Britain, *Fiscal Studies*, vol.26, no.1, pp. 35-53

Burstrom, B., and Fredlund, P., (2001), Self-rated Health: Is It as Good a Predictor of Subsequent Mortality Among Adults in Lower as Well as in Higher Social Classes? *Epidemiology and Community Health*, 55(11), pp. 836–840.

Case, A. and Paxon, C., (2005), Sex Differences in Morbidity and Mortality, *Demography*, vol. 42, no. 2, pp. 189-214

Connelly, J.E., Philbrick, J.T., Smith, R.G., Kaiser, D.L., and Wymer, A., (1989), Health Perceptions of Primary Care Patients and the Influence on Health Care Utilisation, *Medical Care* 27, S99–S109.

DeSalvo, K, et al., (2005), Predicting Mortality and Healthcare Utilization with a Single Question, *Health Services Research*, vol. 40, no. 4, August 2005, pp. 1234-46

Finkelstein, A., and Poterba, J., (2002), Selection Effects in the United Kingdom Individual Annuities Market, *The Economic Journal*, 112, pp. 28-42

Finkelstein, A., and Poterba, J., (2004), Adverse Selection in Insurance Markets: Policyholder Evidence from the U.K. Annuity Market, *The Journal of Political Economy[Chicago]*, vol.112, pt.1, no.1, pp.183-208

Finkelstein, A., and Poterba, J., (2006), Testing for Adverse Selection with "Unused Observables", NBER Working Paper No. 12112

Hurd, M., and McGarry, K., (2002), The Predictive Validity of Subjective Probabilities of Survival, *Economic Journal*, vol. 112, no. 482, pp. 966-85

Idler, E.L, and Benyamini, Y, (1997), Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies**, *Journal of Health and Social Behavior*, Vol. 38, No. 1, pp. 21-37

Idler, E.L., and Kasl, S.V., (1995), Self-Ratings of Health: Do They Also Predict Change in Functional Ability?, *Journal of Gerontology* 50B, S344–S353

James, E., Vittas, D., (1999), Annuity Markets in Comparative Perspective: Do Consumers Get Their Money's Worth?, The World Bank, Policy Research Working Paper Series: 2493

McCallum, J., Shadbolt, B., and Wang, D., (1994), Self-Rated Health and Survival: 7 Years Follow-up Study of Australian Elderly, *American Journal of Public Health* 847, pp. 1100–1105

Milevsky, M., (1998), Optimal Asset Allocation Toward the End of the Lifecycle: To Annuitize or Not to Annuitize?, *The Journal of Risk and Insurance*, 65(3), pp. 401-426

Mitchell, O., Poterba, J., and M. J. Warshawsky, (1997), New Evidence on the Money's Worth of Individual Annuities, Working paper 6002, National Bureau of Economic Research, Inc., Cambridge, MA

Mitchell,O., Poterba, J., Warshawsky, M., and Brown, J., (1999), New Evidence on Money's Worth of Individual Annuities, *American Economic Review*, 89, pp. 1299-1318

Mossey, JM, and Shapiro, E., (1982), Self-Rated Health: A Predictor of Mortality Among the Elderly, *American Journal of Public Health,* 72, pp. 800-808

Okun, M.A., Stock, W.A., Haring, M.J., and Witter, R.A., (1984), Health and Subjective Well-Being: A Meta-Analysis, *International Journal of Ageing and Human Development* 192, pp. 111–132

Pollock, R., and Suyderhoud, J.P., (1992), An Empirical Window on Rational Expectations Formation, *Review of Economics and Statistics*, vol. 74, no. 2, pp. 320-24

Poterba, J., (2001), Annuity Markets and Retirement Security, *Fiscal Studies*, vol. 22, no. 3, pp. 249-270

Salas, C., (2002), On the Empirical Association between Poor Health and Low Socioeconomic Status at Old Age, *Health Economics*, vol. 11, no. 3, pp. 207-20

Schwarze, J, Andersen, H., and Anger, S., (2000), Self-Rated and Changes in Self-Rated Health as Predictors of Mortality - First Evidence from German Panel Data, Humboldt Universitaet Berlin, Sonderforschungsbereich 373

Smith, J., (1999), Healthy Bodies and Thick Wallets: The Dual Relation Between Health and Economic Status, *Journal of Economic Perspectives*, vol. 13, no. 2, pp. 145-166

Smith, VK, Taylor, DH Jr., and Sloan, FA, (2000), Longevity Expectations and Death: Can People Predict Their Own Demise?, Duke University, Department of Economics, Working Papers: 00-15

Statistics Canada, (1996), NPHS Public Use Microdata Documentation, Statistics Canada

Van Doorslaer, E., and Gerdtham, (2003), Does Inequality in Self-Assessed Health Predict Inequality in Survival by Income?- Evidence from Swedish Data, *Social Science and Medicine*, 57, pp.1621-1629

Van Doorslaer, E., and Koolman, X., (2000), Income-Related Inequalities in Health in Europe: Evidence from the European Community Household Panel, Ecuity II Project, Working Paper #1, Erasmus University, Rotterdam

Van Doorslaer, E., Wagstaff, A., Bleichrodt, H., et al, (1997), Income-Related Inequalities in Health: Some International Comparisons, *Journal of Health Economics* 16, pp. 93-112

Van Ourti, T., (2003), Socio-Economic Inequality in Ill-Health Amongst the Elderly. Should One Use Current or Permanent Income?, *Journal of Health Economics*, 22(2), pp. 219-41

Verbeek M, Nijman T., (1992), Testing for selectivity bias in panel data models, *International Economic Review*, vol.33, no.3, pp. 681-703

Walliser, J., (2000), Adverse Selection in the Annuities Market and the Impact of Privatizing Social Security, *Scandinavian Journal of Economics*, vol. 102 no. 3, pp. 373-394

Warshawsky, M., (2001), The Market for Individual Life Annuities and the Reform of Social Security: An Update and Further Analysis, *Benefits Quarterly [Brookfield]*, vol.17, no.4, pp. 24-43

APPENDIX

Table 1. DISTRIBUTION OF SELF-ASSESSED HEALTH (SAH) BY GENDER AND AGE-GROUPS

Sample	SAH	Ages			
		All	20 to 34	35 to 49	50 to 64
Males	Excellent/Very Good	0.70	0.77	0.70	0.60
	Good	0.24	0.20	0.25	0.27
	Fair/Poor	0.06	0.03	0.05	0.13
Total Sample		4516	1677	1733	1106
Females	Excellent/Very Good	0.67	0.73	0.68	0.55
	Good	0.24	0.22	0.24	0.29
	Fair/Poor	0.09	0.05	0.08	0.16
Total Sample		4488	1544	1655	1289

TABLE 2. DESCRIPTION OF FUTURE HEALTH CONDITIONS

Health Condition	Prevalence of Condition	Description
Deceased	Over the past 10 years	1: Individual is deceased within 10 years after the year of initial observation 0: Otherwise
Major Condition	Current	1: Individual has a Major Condition (heart disease, cancer, stroke) 10 years after the year of initial observation 0: Otherwise
	Over the past 10 years	1: Individual has experienced a Major Condition over the 10 years after the year of initial observation 0: Otherwise
Medium Condition	Current	1: Individual has a Medium Condition (diabetes, hypertension) 10 years after the year of initial observation 0: Otherwise
	Over the past 10 years	1: Individual has experienced a Medium Condition over the 10 years after the year of initial observation 0: Otherwise
Restricted (long-term)/ Restricted (LT)	Current	1: Because of a physical or mental condition or a health problem the individual is limited in the kind or amount of activity they can perform at home, school, work or other 10 years after the year of initial observation 0: Otherwise
	Over the past 10 years	1: Because of a physical or mental condition or a health problem the individual has been limited in the kind or amount of activity they can perform at home, school, work or other over the 10 years after the year of initial observation 0: Otherwise
Minor Condition	Current	1: Individual has a Minor Condition (all but major and medium) 10 years after the year of initial observation 0: Otherwise
	Over the past 10 years	1: Individual has experienced a Minor Condition over the 10 years after the year of initial observation 0: Otherwise
Restricted (short-term)	Current	1: Because of a physical or mental condition or a health problem the individual is limited in the kind or amount of activity they can perform at home, school, work or other (for a period less than 6 months) 10 years after the year of initial observation 0: Otherwise
	Over the past 10 years	1: Because of a physical or mental condition or a health problem the individual has been limited in the kind or amount of activity they can perform at home, school, work or other (for a period less than 6 months) over the 10 years after the year of initial observation 0: Otherwise

Table 3. PREVALENCE RATES OF HEALTH CONDITIONS

Condition	Prevalence of Condition	Sample			
		Male		Female	
		Cycle 1	Cycle 5	Cycle 1	Cycle 5
Deceased	Over the past 10 years		0.05		0.03
Major Condition	Current	0.02	0.07	0.03	0.05
	Over the past 10 years		0.10		0.10
Medium Condition	Current	0.07	0.15	0.07	0.17
	Over the past 10 years		0.19		0.19
Restricted (LT)	Current	0.12	0.15	0.11	0.13
	Over the past 10 years		0.26		0.25
Minor Condition	Current	0.44	0.56	0.50	0.66
	Over the past 10 years		0.76		0.82
Restricted (short-term)	Current	0.11	0.16	0.14	0.18
	Over the past 10 years		0.29		0.33

Notes:
1. Current indicates current prevalence of a condition
2. Over the past 10 years spans the period from Cycle 1 to Cycle 5 and indicated prevalence over those 10 years. The condition could also be currently existent

Table 4. SUMMARY STATISTICS – SOCIO-ECONOMIC VARIABLES

Variable	Mean	Standard Deviation
Mother Tongue: French	0.27	0.44
Mother Tongue: Other	0.15	0.36
Immigrant	0.19	0.39
Age	39.61	11.63
Household Income: $30,000-$49,000	0.30	0.46
Household Income: $50,000-$79,000	0.28	0.45
Household Income: $80,000 or over	0.15	0.36
Secondary School Graduate	0.17	0.37
Post-secondary Certificate	0.27	0.45
College or University Education	0.39	0.49
Married/Common Law	0.72	0.45
Male	0.51	0.50
Smoker	0.33	0.47
Drinker	0.84	0.37
Body Mass Index	24.58	4.30
Full-time Employee	0.64	0.48
Part-time Employee	0.10	0.30
Unemployed	0.05	0.21
Self-employed	0.11	0.32
Residence: Quebec	0.26	0.44
Residence: Ontario	0.37	0.48
Residence: Prairies	0.16	0.37
Residence: British Columbia	0.13	0.33

**Table 5. BASELINE RISKS BY GENDER, AGE-GROUPS AND
HEALTH CONDITIONS, LOGIT MODEL**

Condition	Sample	Ages			
		All	20 to 34	35 to 49	50 to 64
Deceased	Male	0.044	0.013	0.022	0.118
	Female	0.024	0.006	0.017	0.064
Major Condition	Male	0.084	0.015	0.068	0.204
	Female	0.060	0.022	0.043	0.136
Medium Condition	Male	0.177	0.048	0.179	0.371
	Female	0.210	0.066	0.189	0.449
Restricted (LT)	Male	0.169	0.112	0.177	0.242
	Female	0.162	0.096	0.186	0.224

Notes:
1. Baseline risk is the probability that a person reporting good SAH experiences a particular health condition. Risks are estimated based on a logit specification.

**Table 6. MARGINAL EFFECTS OF SELF-ASSESSED HEALTH (SAH)
MALES, BY AGE-GROUPS AND HEALTH CONDITIONS,
(LOGIT MODELS)**

Condition	SAH	Males of Age			
		All	20 to 34	35 to 49	50 to 64
Deceased	Excellent/ Very Good	-0.015** (0.007)	0.009 (0.007)	-0.005 (0.009)	-0.060*** (0.022)
	Fair/Poor	0.010 (0.009)	0.048 (0.048)	0.042* (0.023)	-0.003 (0.026)
Major Condition	Excellent/ Very Good	-0.0002 (0.010)	0.031*** (0.010)	0.010 (0.015)	-0.040 (0.033)
	Fair/Poor	0.024 (0.017)	0.424*** (0.160)	0.008 (0.027)	0.021 (0.042)
Medium Condition	Excellent/ Very Good	-0.049*** (0.014)	-0.018 (0.015)	-0.060*** (0.023)	-0.059 (0.037)
	Fair/Poor	-0.023 (0.019)	-0.017 (0.017)	-0.033 (0.038)	-0.032 (0.047)
Restricted (LT)	Excellent/ Very Good	-0.048*** (0.015)	-0.026 (0.021)	-0.039 (0.024)	-0.079** (0.035)
	Fair/Poor	0.040* (0.024)	0.036 (0.041)	0.102** (0.052)	0.036 (0.045)

Notes:
1. The marginal effect of a dummy variable is the change in the probability of the outcome for a discrete change of the dummy from 0 to 1.
2. Effects are relative to the base category, which is "good" self-assessed health
3. Standard errors are in parentheses
4. Significance levels: *** p<0.01, ** p<0.05, * p<0.1

Table 6A. TESTS OF EQUALITY OF MARGINAL EFFECTS ACROSS PAIRS OF AGE-GROUPS

P-VALUES (5%)

Males, by Health Condition
Marginal Effects of SAH

		Age-group	
Condition	Age-group	20 to 34	35 to 49
Marginal Effect of Excellent/Very Good (versus Good)			
Deceased	35 to 49	0.215	
	50 to 64	0.003	0.021
Major Condition	35 to 49	0.245	
	50 to 64	0.040	0.168
Medium Condition	35 to 49	0.125	
	50 to 64	0.304	0.980
Restricted (LT)	35 to 49	0.693	
	50 to 64	0.194	0.341
Marginal Effect of Fair/Poor (versus Good)			
Deceased	35 to 49	0.906	
	50 to 64	0.351	0.199
Major Condition	35 to 49	0.010	
	50 to 64	0.015	0.785
Medium Condition	35 to 49	0.701	
	50 to 64	0.766	0.985
Restricted (LT)	35 to 49	0.320	
	50 to 64	0.992	0.334

Table 7. MARGINAL EFFECTS OF SELF-ASSESSED HEALTH (SAH)
FEMALES, BY AGE-GROUPS AND HEALTH CONDITIONS,
(LOGIT MODELS)

Condition	SAH	Females of Age			
		All	20 to 34	35 to 49	50 to 64
Deceased	Excellent/ Very Good	0.006 (0.006)	0.001 (0.005)	0.009 (0.008)	0.008 (0.018)
	Fair/Poor	0.012 (0.008)	0.004 (0.015)	0.010 (0.012)	0.020 (0.020)
Major Condition	Excellent/ Very Good	-0.030*** (0.009)	-0.007 (0.010)	-0.022* (0.013)	-0.067*** (0.025)
	Fair/Poor	0.001 (0.009)	0.021 (0.021)	-0.006 (0.014)	0.013 (0.026)
Medium Condition	Excellent/ Very Good	0.011 (0.012)	0.018 (0.013)	-0.022 (0.021)	0.059* (0.031)
	Fair/Poor	0.043* (0.022)	-0.006 (0.034)	0.042 (0.038)	0.113** (0.050)
Restricted (LT)	Excellent/ Very Good	-0.039*** (0.013)	-0.030 (0.018)	-0.035 (0.023)	-0.071** (0.030)
	Fair/Poor	0.090*** (0.021)	0.050 (0.033)	0.175*** (0.045)	0.026 (0.034)

Notes:
1. The marginal effect of a dummy variable is the change in the probability of the outcome for a discrete change of the dummy from 0 to 1.
2. Effects are relative to the base category, which is "good" self-assessed health
3. Standard errors are in parentheses
4. Significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 7A. TESTS OF EQUALITY OF MARGINAL EFFECTS ACROSS PAIRS OF AGE-GROUPS

P-VALUES (5%)

Females, by Health Condition
Marginal Effects of SAH

Condition	Age-group	Age-group 20 to 34	35 to 49
Marginal Effect of Excellent/Very Good (versus Good)			
Deceased	35 to 49	0.372	
	50 to 64	0.728	0.933
Major Condition	35 to 49	0.346	
	50 to 64	0.026	0.114
Medium Condition	35 to 49	0.099	
	50 to 64	0.225	0.030
Restricted (LT)	35 to 49	0.761	
	50 to 64	0.202	0.345
Marginal Effect of Fair/Poor (versus Good)			
Deceased	35 to 49	0.726	
	50 to 64	0.521	0.690
Major Condition	35 to 49	0.291	
	50 to 64	0.813	0.526
Medium Condition	35 to 49	0.349	
	50 to 64	0.049	0.256
Restricted (LT)	35 to 49	0.025	
	50 to 64	0.623	0.008

**Table 8. ODDS-RATIOS FOR SELF-ASSESSED HEALTH FOR
MALES, BY AGE-GROUPS AND HEALTH CONDITIONS,
(LOGIT MODEL)**

Condition	SAH	Males of Age			
		All	20 to 34	35 to 49	50 to 64
Deceased	Excellent/ Very Good	0.66** (0.45 - 0.95)	2.44 (0.53 - 11.22)	0.79 (0.34 - 1.83)	0.51*** (0.30 - 0.84)
	Fair/Poor	1.30 (0.83 - 2.04)	6.62* (0.83 - 52.65)	3.88*** (1.49 - 10.04)	0.97 (0.55 - 1.72)
Major Condition	Excellent/ Very Good	0.10 (0.71 - 1.40)	27.85*** (2.94 - 263.92)	1.25 (0.67 - 2.33)	0.74 (0.46 - 1.20)
	Fair/Poor	1.47 (0.91 - 2.36)	346.40*** (17.8 - 6746.1)	1.16 (0.43 - 3.12)	1.17 (0.64 - 2.14)
Medium Condition	Excellent/ Very Good	0.61*** (0.47 - 0.79)	0.60 (0.28 - 1.28)	0.58*** (0.40 - 0.84)	0.71 (0.47 - 1.08)
	Fair/Poor	0.78 (0.50 - 1.20)	0.55 (0.13 - 2.34)	0.71 (0.30 - 1.68)	0.83 (0.46 - 1.46)
Restricted (LT)	Excellent/ Very Good	0.66*** (0.52 - 0.84)	0.74 (0.46 - 1.17)	0.72* (0.49 - 1.05)	0.58** (0.37 - 0.92)
	Fair/Poor	1.39* (0.97 - 1.99)	1.49 (0.68 - 3.24)	2.11** (1.12 - 3.97)	1.28 (0.72 - 2.27)

Notes:
1. 95% confidence interval is reported in parentheses
2. Effects are relative to the base category, which is "good" self-assessed health
3. Significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 9. ODDS-RATIOS FOR SELF-ASSESSED HEALTH FOR FEMALES, BY AGE-GROUPS AND HEALTH CONDITIONS, (LOGIT MODEL)

Condition	SAH	Females of Age			
		All	20 to 34	35 to 49	50 to 64
Deceased	Excellent/ Very Good	1.33 (0.81 - 2.20)	1.20 (0.19 - 7.73)	1.79 (0.74 - 4.34)	1.15 (0.60 - 2.20)
	Fair/Poor	1.65* (0.95 - 2.84)	1.64 (0.06 - 45.471)	1.79 (0.55 - 5.83)	1.41 (0.73 - 2.71)
Major Condition	Excellent/ Very Good	0.53*** (0.38 - 0.76)	0.71 (0.28 - 1.82)	0.55* (0.29 - 1.06)	0.51*** (0.30 - 0.84)
	Fair/Poor	1.03 (0.70 - 1.51)	2.32 (0.67 - 8.05)	0.84 (0.36 - 1.96)	1.14 (0.69 - 1.87)
Medium Condition	Excellent/ Very Good	1.12 (0.88 - 1.42)	1.46 (0.83 - 2.57)	0.81 (0.55 - 1.19)	1.45* (0.99 - 2.13)
	Fair/Poor	1.47** (1.02 - 2.13)	0.89 (0.22 - 3.63)	1.45 (0.79 - 2.67)	1.93** (1.11 - 3.32)
Restricted (LT)	Excellent/ Very Good	0.70*** (0.56 - 0.88)	0.70 (0.45 - 1.11)	0.74 (0.51 - 1.07)	0.60** (0.40 - 0.91)
	Fair/Poor	2.05*** (1.54 - 2.73)	1.80* (0.94 - 3.44)	3.28*** (2.03 - 5.31)	1.21 (0.76 - 1.93)

Notes:
1. 95% confidence interval is reported in parentheses
2. Effects are relative to the base category, which is "good" self-assessed health
3. Significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$

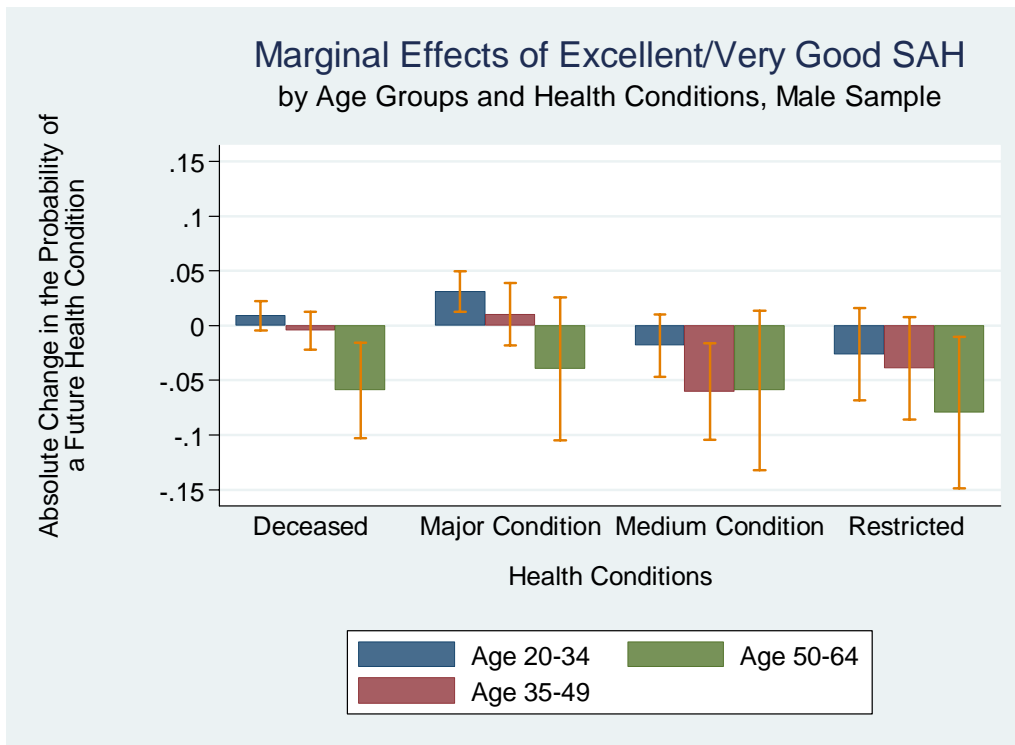**Figure 1. Male Sample, Marginal Effects of Excellent/Very Good SAH by Age Groups and Health Conditions**

Marginal Effects of Excellent/Very Good SAH
by Age Groups and Health Conditions, Male Sample

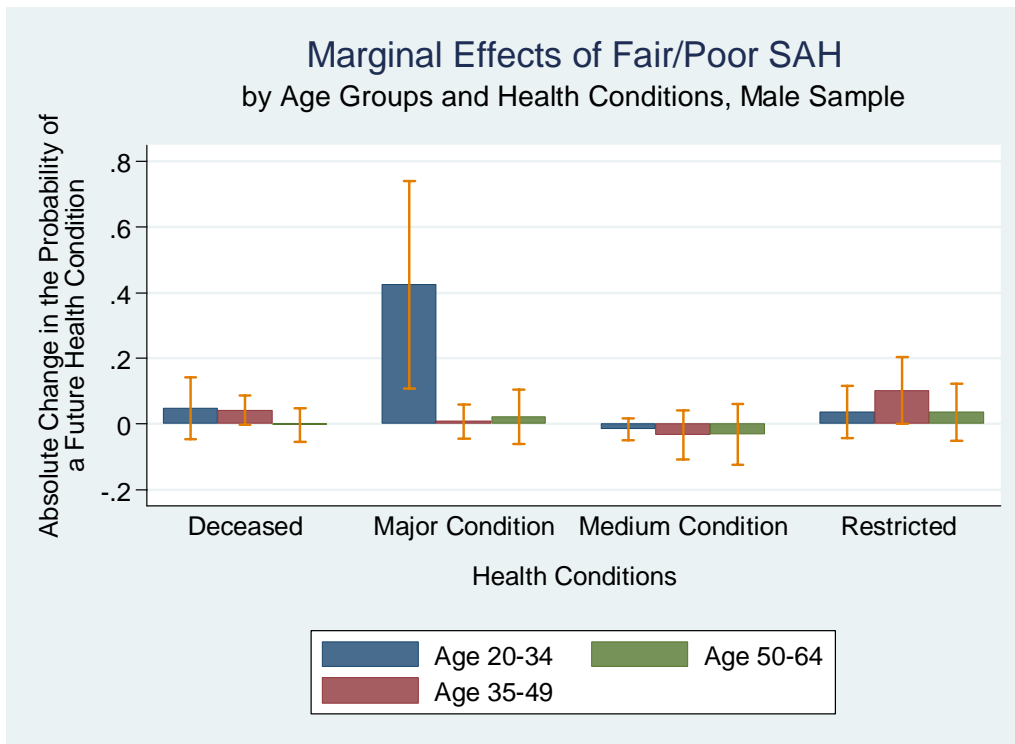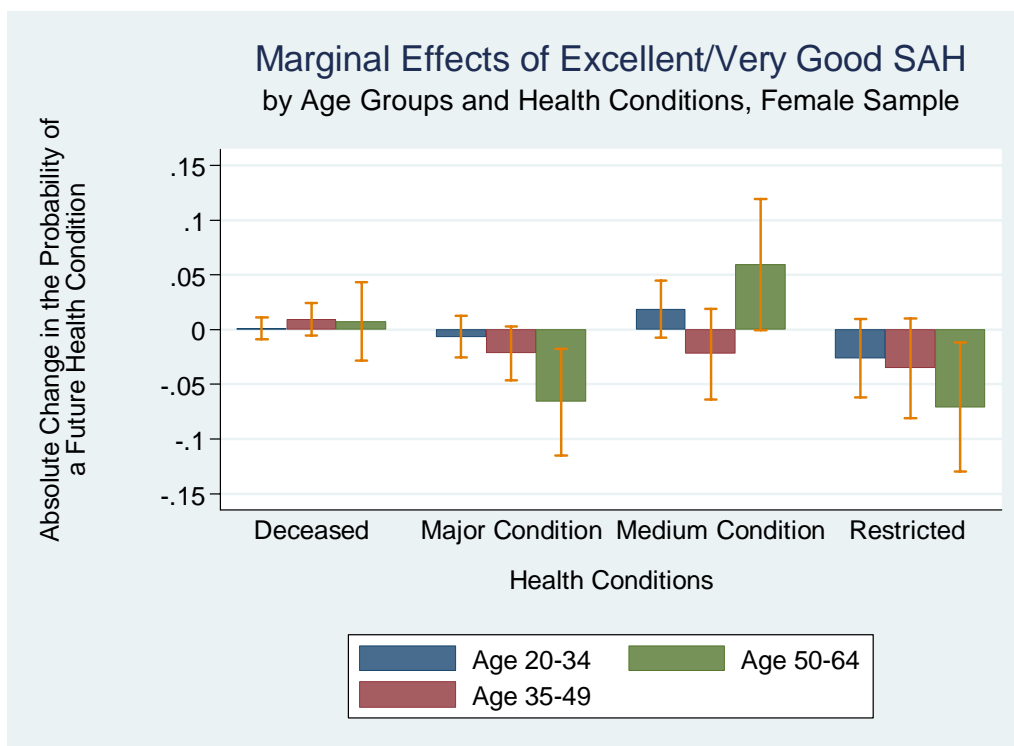**Figure 2. Male Sample, Marginal Effects of Fair/Poor SAH by Age Groups and Health Conditions**

Marginal Effects of Fair/Poor SAH
by Age Groups and Health Conditions, Male Sample

Figure 3. Female Sample, Marginal Effects of Excellent/ Very Good SAH by Age Groups and Health Conditions

Figure 4. Female Sample, Marginal Effects of Fair/Poor SAH by Age Groups and Health Conditions
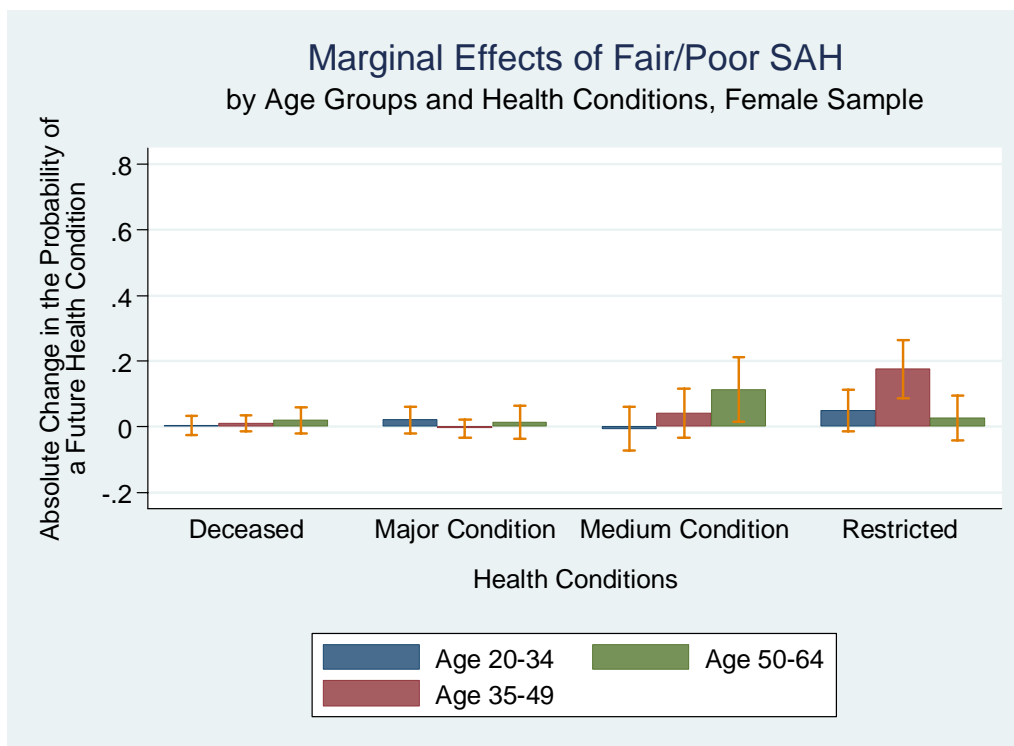
Figure 5. Male Sample, Marginal Effects of Excellent/Very Good SAH by Age Groups, Health Conditions and Horizon
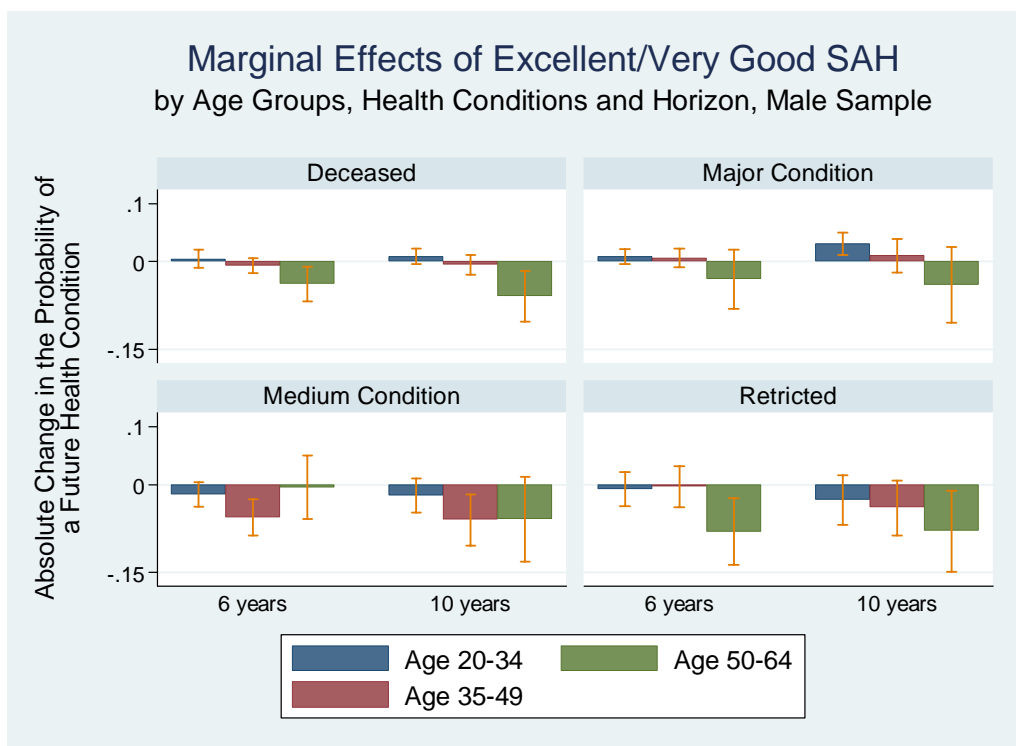

Figure 6. Female Sample, Marginal Effects of Excellent/Very Good SAH by Age Groups, Health Conditions and Horizon