NBER WORKING PAPER SERIES

WHAT DO SELF-REPORTED, OBJECTIVE, MEASURES OF HEALTH MEASURE?

Michael Baker
Mark Stabile
Catherine Deri

Working Paper 8419
http://www.nber.org/papers/w8419

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2001

Baker gratefully acknowledges the research support of SSHRC (Grant # 410-99-0112). Stabile and Deri gratefully acknowledge support from the Institute for Clinical and Evaluative Sciences. We would like to thank John Ham, David Cutler and seminar participants at various workshops for helpful comments. The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research.

What do Self-Reported, Objective, Measures of Health Measure?
Michael Baker, Mark Stabile and Catherine Deri
NBER Working Paper No. 8419
August 2001
JEL No. I12, J22

ABSTRACT

Survey reports of the incidence of chronic conditions are considered by many researchers to be more objective, and thus preferable, measures of unobserved health status than self-assessed measures of global well being. The former are 1) responses to specific questions about different ailments, which may constrain the likelihood that respondents rationalize their own behavior through their answers, and 2) more comparable across respondents. In this paper we evaluate this hypothesis by exploring measurement error in these "objective, self-reported" measures of health. Our analysis makes use of a unique data set that matches a variety of self-reports of health with respondents' medical records. Our findings are striking. For example, the ratio of the error variance to the total variance ranges from just over 30 percent for the incidence of diabetes to over 80 percent for the incidence of arthritis. Furthermore, for many conditions the error is significantly related to individuals' labor market activity, as hypothesized in the literature. In the final section of the paper we compare estimates of the effect of these different measures of health on labor market activity.

Michael Baker
Department of Economics
University of Toronto
150 St. George Street
Toronto, ON
Canada M5S 3G7
and NBER
baker@chass.utoronto.ca

Mark Stabile
Department of Economics
University of Toronto
150 St. George Street
Toronto, ON
Canada M5S 3G7
and NBER
mark.stabile@utoronto.ca

Catherine Deri
Institute for Policy Analysis
University of Toronto
140 St. George Street
Toronto, ON
Canada M5S 3G6
cderi@chass.utoronto.ca

Introduction

The limitations of using subjective, self reported assessments of global health and/or physical capacity as explanatory variables in empirical models of labor market behavior are now widely recognized in the literature (see Currie and Madrian 1999 for a review). The problems range from the so called "justification hypothesis"—health problems are a socially acceptable and convenient rationalization of absence from the labor market—to the unknown level of comparability of these subjective evaluations across individuals. In response, many researchers put greater stock in more objective, but still self-reported, measures of specific illnesses or information on subsequent mortality as proxies for health (Bound and Burkhauser 1999). The argument here is that the specificity of the questions constrains the likelihood that respondents rationalize their own behavior through their answers.

As argued by Bound (1991) among others however, these alternative measures are not a panacea. First, self-reported global health measures are subject to both negative and positive biases—resulting from measurement error and endogeneity respectively—that arguably cancel each other out. If more objective measures are less prone to the rationalization bias, then they are only subject to the attenuation bias resulting from measurement error. Second, instrumental variable (IV) strategies of using the more objective measures as instruments for self reported health may retrieve less biased estimates of the effects of health on the labor market outcome of interest, but may lead to biased estimates of the parameters of any other explanatory variables that are themselves determinants of self-reported health.

In this paper we investigate the validity of self-reported measures of specific ailments. Our starting point is the 1994 Canadian National Population Health Survey (NPHS), which is a nationally representative survey of health including measures of 1) self-reported global health, 2) specific work and activity limitations and 3) the self-reported incidence of specific ailments. For respondents in the province of Ontario, these data have been matched with diagnosis/treatment information (following the International Classification of Diseases standard – 9^{th} revision (ICD-9)) taken from the Ontario Health Insurance Plan (OHIP) for the survey and two preceding years. OHIP is a public health care program financed out of tax revenues, which covers all individuals subject to certain residency requirements. Since private alternatives to public health are either prohibited by law or are relatively very expensive (e.g., going to US health care providers), these OHIP health records should provide a very complete record of the diagnoses and treatments of these individuals over the period.[1]

The diagnoses/treatment information contained in the OHIP records provides a unique opportunity to validate the self-reports of diseases taken from the NPHS. On one hand, the NPHS question is fairly precise: "[Do you]…have any of the following long-term conditions that have been diagnosed by a health professional"? A list of possible ailments is offered to respondents, thus avoiding the under-reporting that might result from reliance on free recall.[2] The meanings of terms used in the question, such as "long-term", are carefully explained. The resulting measures, therefore, exemplify the specificity that is associated with objective, self-reported health variables. On the other hand, the records of diagnoses and treatment taken from the OHIP records are arguably the "truth" that the NPHS questions attempt to record. A remaining issue, of course, is the relation between either of these measures and an individual's health defined as work capacity, which both presumably measure with error.

[1] The 1996 NPHS asked the respondents whether that had received health care in the United States in the past 12 months. Fewer than 1 percent of the sample (the sample is Canada wide) reported receiving such care.

[2] See Bound, Brown and Mathiowetz (2000) for a discussion of some of the literature on measurement error in health reports based on free recall versus recognition.

We begin by looking at the relationship between subjective and objective self-reports of health, and provide a comparison of these NPHS variables to similar measures retrieved from the US Health and Retirement Survey (HRS) to provide some context for our findings. Next we investigate the measurement error in the NPHS variables. This includes summaries of the incidence of "false positive" and "false negative" reporting for the major disease categories. Our findings here are striking. For many of the diseases more than 50 percent of the individuals who have a positive doctor's diagnosis in the OHIP data fail to report having the disease in the NPHS data. Similarly, we frequently find that more than 50 percent of individuals who report having diseases in the NPHS have no corresponding doctor's diagnosis in the OHIP records. A series of exercises are conducted to attempt to bound the measurement error and show that while there is likely some error in the OHIP data the significant error that we find comes from the NPHS.

We then quantify the measurement error using tools familiar from studies of measurement error in labor market data. Our estimates of the ratio of the error variance to total variance are often larger than 0.5 and in some cases as large as 0.8. In the following sections we test and find evidence for the "justification hypothesis" and show how the error in self reported health varies with the intensity of the condition.

As a final check on the generality of our inference, we investigate the relationship between work and the different measures of health within a simple model of labor market participation. We compare different estimation strategies used in previous studies, including using objective measures of health as instruments for more subjective measures. Consistent with the previous literature, the IV estimates of the effect of health on participation are uniformly much larger than their OLS counterparts.

The Data

The 1994/95 NPHS is a nationally representative survey of over 23,000 Canadian households conducted in the last three quarters of 1994 and first quarter of 1995. While some minimal information was collected for all members of each household, a randomly selected member, 12 years of age or older, participated in the in-depth interview. The information collected included various measures of health status, use of health services and the presence of various risk factors. A limited amount of demographic and socio-economic information was also collected.

Just over 5000 households were interviewed for the Ontario sample. Individuals completing the in-depth interview were requested to supply their Ontario health number, which is used for all publicly insured services in the province. The claims records for these individuals for the two years prior to their NPHS interview date have been linked to the survey responses by the Ontario Ministry of Health in collaboration with Statistics Canada. Ninety-three percent of the Ontario respondents provided their health card numbers and consented to the linkage of the records. The Ontario Ministry of Health considered 96 percent of these numbers as valid for linkage (Williams et al, 1998). The main elements of the OHIP data are the code for service provided, the date of service, one associated diagnosis and fee paid. Therefore, diagnosis of a new disease and/or treatment of an on-going condition that involved a health care provider that can claim under OHIP should be captured in these data.[3]

[3] Excluded from this OHIP database are claims under the provincial worker's compensation program and remuneration to physicians through Alternative Fee Plans (AFPs). The latter are concentrated in three hospitals located in Kingston, Toronto, Sault Ste Marie and Health Service Organizations located primary in the Hamilton-Wentworth and Waterloo areas. AFPs account for roughly 5 percent of total physician expenditures. We have rerun the measurement error exercises (table 3a) of our analysis excluding the Central West Region of Ontario (i.e. excluding Hamilton-Wentworth and Waterloo areas) and found no significant difference in our results. These results are available on request. We are currently gathering information on the incidence of Worker's Compensation claims by ailment in an attempt to determine the bias that might result from the exclusion of these claims.

All of the 4621 Ontario observations from the NPHS that were valid for linkage to the respondent's OHIP records are used in the majority of our analysis. In analysis focusing on the relationship between health and labor supply decisions, however, we exclude those under age 16 and those going to school.

Some general characteristics of the sample are reported in Table 1. We assess the accuracy of the reported incidence of the 13 ailments listed here. These ailments, representing 65 percent of the chronic conditions captured in the NPHS, were selected to be comparable with conditions used in previous studies of self-reported health.[4] The conditions that were not included are food allergies, other allergies, epilepsy, acne, Alzheimer's disease, heart disease and urinary incontinence.

To relate these reports to the OHIP records, the chronic conditions were mapped into ICD-9 codes. This process is not completely straightforward, as neither is the relationship between the ailments and the codes self evident, nor is which codes get reported as a specific condition necessarily stable across health care providers. As a result, we constructed two mappings that alternatively attempt to take a broad and very precise view of the relationship. The algorithms are reported in Appendix A.

An important question is whether these OHIP codes actually represent the object that the NPHS attempts to measure? As noted in the introduction, the NPHS question investigates the current experience of a long-term condition as diagnosed by a health professional. Respondents are reminded that "long-term" ailments last 6 months or more. Certainly the insistence that the ailment has been diagnosed by a professional suggests that the OHIP data should capture the event. Although professionals are not explicitly defined at this point in the interview, at an earlier stage individuals are questioned about their interactions with professionals over the last 12 months, including general practitioners and family physicians, eye specialists, other medical doctors or nurses, dentists, chiropractors, physiotherapists, social workers, psychologists and speech, audiology or occupational therapists. The services of all of these professionals except for dentists, social workers and speech therapists are covered by OHIP. The phrasing and structure of the questions also draws a distinction between health professionals and alternative health providers, such as acupuncturists and homeopathic medical doctors, who are also not covered by the OHIP program. Our focus on medical conditions makes it unlikely that these exclusions in OHIP coverage lead to the mismatches between the NPHS variables and the OHIP codes that we uncover.

Another issue is whether the 2-year window for the OHIP codes is sufficient to capture diagnoses/treatments underlying the responses to the NPHS question. It seems reasonable that "current" ailments are likely to have been diagnosed or treated in the past 2 years, although this will clearly vary with the disease. For example, individuals reporting current incidence of cancer will presumably have had some treatment in this period. Less certain will be a condition such as arthritis, which may have been diagnosed years ago and is not currently receiving active treatment. More generally, we can compare the error in the NPHS responses by disease to discover if any patterns are consistent with these arguments.

Previous Literature

While there is a reasonably large literature investigating the correspondence between individuals' responses to questions about their general health or use of health services, and medical records or physician's appraisals, there are relatively few previous studies of the reliability of self-reports of chronic conditions. Harlow and Linet (1989) survey the literature

---

[4] See, for example, Dwyer and Mitchell (1999)

and uncover only six studies that attempt to validate this sort of data. A common feature of this research is that it involves relatively unrepresentative samples. For example, enrollees in a single American health insurance organization serve as the basis of more than one of these papers. Presumably the selection of individuals into these plans limits the universality of any inference gained from these samples. Other studies focus on individuals in a particular city or area, and/or are based on quite small samples. A clear advantage here is that because there is a single and universal health insurance plan in Ontario, and the NPHS is a representative sample of the province, our inference is much more likely to provide general insights to the problems of measurement error in these sorts of variables.

As noted by Harlow and Linet (1989), some of these studies are also limited by the strategies used to analyze the data. For example, some papers condition inference on the report of an ailment in a medical record, thus ignoring error arising from false positives. Alternatively, others condition on respondents reports of ailments, thus omitting false negatives.

Two statistics commonly reported in these studies are the percentage of medical records of a disease matched with a survey report, and the percentage of interview reports of disease matched with a medical record. These have been interpreted as measures of the degree of under- and over-reporting of specific ailments, respectively. Comparison of these statistics across studies or to similar statistics constructed from the NPHS data is problematic, however, due to cross-study differences in the (ICD) coding of specific diseases and/or in the questions asked of respondents. For example, we take a broader view of "ulcers" than does Madow (1973), including esophageal and other peptic ulcers in addition to stomach and duodenal ulcers in the category. Our coding was chosen to be consistent with the question asked in the NPHS; presumably the survey instrument in the Madow study allowed some distinction between different types of ulcers. Strictly speaking our results will be comparable to Madow's if there is no heterogeneity in any measurement error by type of ulcer; that is, if this is true we can attribute any dissimilarity between our and Madow's results to differences in the representativeness of the samples rather than differences in the objects being analyzed.

With these qualifications in mind, in Appendix B we present these measures of under-and over-reporting of chronic conditions for the NPHS sample and compare the results (where possible) to those in previous studies. For some diseases in some studies there is a fair bit of agreement with the results from the NPHS, while in other cases very little. There is, perhaps, somewhat greater accord in the relative levels of the statistics across ailments. Conditions such as bronchitis and sinusitis would appear to be more prone to over-reporting, while cancer appears to be more prone to under-reporting.

There is also one previous study of our NPHS data. Hux et al. (2001) match individuals in the NPHS to various diabetes databases for Ontario. Matching registered diabetics to the survey responses, they find considerable under-reporting of diabetes in the NPHS. Their results are consistent with the measurement error we report for this condition below.

An Overview of Self-Reported Subjective and Objective Measures of Health

While subjective and objective self-reports of health status are intended to measure different objects, it is of interest to discover their relationship since ultimately we are interested in the latent variable work capacity for which either might serve as a proxy. In Figures 1a through 1m, we provide an overview of this relationship by graphing the distribution of responses to a self-report of global health by the self-reported incidence of our 13 diseases. The global health question is fairly standard asking "How would you consider your health?", and allowing for the possible answers: Excellent, Very Good, Good, Fair and Poor.

The general pattern in the figures is that individuals with a given disease have a similar shaped distribution to those who don't except that it is shifted to the right: that is towards poorer health. For example, of the 259 respondents with a diagnosis of cancer, 13 percent consider themselves to be in excellent health, 29 percent in very good, 32 percent in good health, 17 percent in fair health and 9 percent in poor health. In comparison, among individuals without a diagnosis of cancer, 25 percent consider themselves in excellent health, 38 percent in very good, 25 percent in good, 9 percent in fair and 3 percent in poor health. Not surprisingly, for many of the less severe conditions, such as migraines and sinusitis, the two distributions are almost indistinguishable.

The pattern holds up for all the ailments with the possible exception of strokes. The 'without strokes' group has as its mode at very good health while the 'with stroke' group has it mode at fair health. This is not unexpected as strokes can be life threatening and can have long lasting/health limiting consequences. However, 16 percent of the respondents who had strokes considered themselves to be in excellent health. This is perhaps surprising. One possibility is that the stroke was sometime ago and the individual has since adopted a healthy lifestyle and is in better health than prior to the stroke. That said, the NPHS question demands that the individual currently be experiencing the long-term consequences of the stroke, which would seem to suggest that health should in some sense have deteriorated. Another possibility is that the individual's point of reference has changed, so that what is reported is that these individuals are in excellent health given that they have had a stroke. If true, this would be an example of how subjective measures can be less than strictly comparable across individuals.

Thus it seems that having any of these conditions translates into, on average, poorer self-assessed health, although no one condition makes all people report that they are in the poorest of health categories. Are the differences statistically significant? We have run ordered probits of the global health measure on each of the self-reported health diagnoses as well as on education, sex, a quartic in age, and marital status. The results suggest that all the diseases we investigate, with the exception of cateracts, are significantly correlated with self-reporting worse health.[5]

To attempt to calibrate the NPHS responses to measures of self reported health available in other data sets we compare these distributions of self-assessed health to comparable information taken from HRS which is increasingly used for research on the impact of health on retirement in the United States. Because the HRS is focused on older individuals, we construct an NPHS sample which is similar in age to the 1992 HRS by restricting the NPHS sample to individuals aged 51-61[6].

It appears that the distribution of self-assessed health is similar between the two samples. In figures 2a-2f we provide some representative inference for cancer, back problems and hypertension. Our results suggest that for individuals without self-reported chronic conditions, the distribution of self-assessed health is very similar. For each ailment the distributions in the NPHS and HRS are almost identical. In the second column we compare the distributions for individuals who do self-report these three conditions. The distributions from the NPHS and HRS are greatly similar, although in the case of cancer, a higher proportion of individuals in the HRS report worse health than in the NPHS. Overall, our comparisons between the NPHS and HRS lead us to conclude that the two data sets are sufficiently alike that the findings we report here should be informative to researchers using HRS data.

---

[5] These results are available from the authors on request.

Measurement Error in Objective Measures of Health

We begin with the simple summary provided by the correlation coefficient between respondents' answers to the objective health questions and the presence of a diagnosis code in the OHIP administrative data for that particular disease. For example, in the case of cancer we calculate the correlation between whether the individual indicated they currently had cancer and whether there is any record of a physician diagnosing them with cancer in the OHIP administrative data. These results are reported in Table 2. The correlations between the interview questions and the diagnoses are much lower than one might expect. For example, the correlation coefficient for cancer is 0.47, for a stroke is 0.48 and for back problems is 0.23. The lowest is for sinusitis, which has a coefficient of 0.13 while the highest is for diabetes, which has a coefficient of 0.71. The coefficient is over 0.50 for only 3 of the 13 conditions.

These low correlation coefficients could result from two different types of errors. First, people may report having a problem that doesn't appear in the administrative data (false positives). Second, people diagnosed with a health problem may not reveal it in the survey (false negatives). Using the "narrow mapping" of diagnosis codes, in table 3a we report the incidence of false positives and negatives in our data (columns 1 and 2), as well as the proportion of false positives as a fraction of the number of people who report having the disease (column 3).

For cancer, 70 percent of those who have been diagnosed with cancer in the past two years do not claim that they have cancer (false negative). Of the people who have no diagnoses of cancer in the past two years, half of a percent claimed that they had cancer (false positive). For stroke, 53 percent of those diagnosed with having a stroke in the past two years do not claim to

---

[6] We do not use the 1994 HRS survey as a comparison group because the questions asked in that round referred to changes in chronic conditions since the previous round (1992) and hence the questions are not directly comparable.

have had a stroke. Again, 0.5 of a percent of people who are not diagnosed with a stroke did claim to have had a stroke. For back problems, 55 percent of those diagnosed with back problems claimed that they did not have a back problem, and a significantly larger 15 percent of those without back problem diagnoses claimed that they did have a back problem. A larger incidence of false positives also appears for other health conditions that one might consider to be more subjective, or more likely self diagnosed. For example, 7 percent of people without a diagnosed migraine claimed to have migraines. For both back problems and migraines false positives play a particularly important role, as they represent 77 percent and 80 percent respectively of all the people who claimed that they had been diagnosed with such problem (column three).

False positives can arise when 1) the individuals are not telling the truth about their condition, 2) they self diagnosed their ailment, 3) they misunderstand their ailment and self-report it as another, 4) they are telling the truth, but were last seen by a doctor for this condition more than two years prior to the interview year (recall that the OHIP data goes back two years), 5) the physician recorded the wrong diagnosis code or the diagnosis code for a second condition that the patient presented was used for billing purposes, or 6) the respondents recently moved to Ontario (the respondents correctly reported the condition but was diagnosed by a physician in their former province or country). The first three accounts are measurement error in the self-reported NPHS variable, while the last three are measurement error in the OHIP records. False negatives can arise if 1) individuals do not feel comfortable reporting their conditions, 2) individuals have very short memories, 3) the individual were cured in the period between the date on which the treatment occurred and when the survey was conducted, or 4) the doctor makes a preliminary diagnosis of the condition that ultimately proves to be false. Here the first two accounts are measurement error in the NPHS variable while the second two are measurement error in the OHIP records.

Of particular concern here is any measurement error in the OHIP data because it undermines our validation exercise. We would expect that for chronic conditions that are more severe in nature such as cancer, diabetes or stroke, the potential for error on the side of the physician is considerably reduced. In these cases, interaction with the health care system is required for treatment and hence the likelihood that OHIP records would not accurately record the condition is low. That we see such large errors in both these types of conditions and less severe conditions suggests that much of the error is indeed in the NPHS and not the OHIP data. Nevertheless, we conduct a number of exercises in an attempt to bound the potential contribution of any OHIP based error to our results.

First consider the fifth account of false positives—the doctor simply enters the wrong diagnosis code when s/he is entering his/her billings. Perhaps s/he enters code 347 instead of 346 for migraines, or perhaps s/he enters a code for a condition very similar to migraines but still within broad classification of diseases of the nervous system and sense organs. One way to discover if this type of error is important is to map the NPHS chronic conditions into the ICD-9 codes in the OHIP data using the “broad mapping” algorithm discussed above (see Appendix A). The broader bands will capture any “small deviations” from the truth in the OHIP data and thus the incidence of false positives should decrease. We use nine major disease categories from the ICD-9 classifications that group diseases into major organs or bodily systems affected. Note that we should also expect the incidence of false negatives to be much higher using these categories, as using broader bands in the OHIP data will capture diseases not investigated in the NPHS data. The results are reported in table 3b. For some ailments the incidence of false positives is virtually unchanged (e.g., cancer, diabetes and migraines) while for others it falls substantially (arthritis, back problems sinus problems). In every case, however, the problem of false positives is not eliminated, and they remain a non-trivial fraction of survey positives. Note also that as expected the incidence of false negatives is now much higher for all ailments. Therefore, there would appear to be a trade off here between significant reductions in false positives for some ailments and significant increases in false negatives for all ailments.

To explore the sixth account of false positives, we repeated the analysis on the sub-sample of respondents who had any type of OHIP claim/record in the previous 2 years. If all the false positives were due to a large number of respondents having just recently moved to Ontario (thus having no OHIP records for our purposes), this exercise should find a sharp decrease in the incidence of false positives. In fact, the decrease in false positives in moving to this sample is very slight, indicating that migration is not driving this sort of error[7,8].

We next address the third account of false negatives: the possibility that respondents were cured of their ailment between the time of diagnosis/treatment and the time of the survey. Given individuals are requested to report the current incidence of long-term condition which must last 6 months or more, limiting the OHIP records to the year directly preceding the interview should reduce the number of false negatives due to this sort of error. The results, reported in table 3c, indicate that, in general, the incidence of false negatives is reduced imposing this condition, the reduction averaging about 10 percent. For many of the ailments the decrease is quite small, however, and for sinus problems the incidence of false negatives actually increases. Note also, that the larger proportionate reductions are for strokes, diabetes, and hypertension. This is a little

[7] The results are not presented as a table but are available from the authors upon request.

[8] Between 1992 and 1994 the average number if people immigrating to Ontario was 67274 from other provinces (CANSIM Label C108224) and 32480 from abroad (CANSIM Label D80). This represents 0.9% of the province's population (CANSIM Label D31241) over this period (Statistics Canada web site: www.statcanada.ca).

surprising since a condition such as chronic diabetes is largely incurable. In addition, when we limit the OHIP data to one year the proportion of false positives increases, as expected. Most importantly, there are still large discrepancies between the self-reports and the administrative data. Therefore, it is unlikely that this sort of error in the OHIP data lies behind the results in table 3a.

The fourth account of false negatives arises if the patient is not properly or completely diagnosed on a first visit to a doctor. For example, suppose a woman goes to a doctor with a lump in her breast and the doctor suspects cancer but isn’t certain. One possibility is that the doctor initially codes the women as having a neoplasm of unspecified or uncertain behavior. This should not precipitate error in the OHIP data, as long as the correct code is entered on subsequent visits once the nature of the lump is determined. Suppose instead, however, that the doctor initially codes the woman with a malignant neoplasm (cancer) on the basis of his/her suspicion. If the lump subsequently turns out to be benign, the OHIP data will record a cancer diagnosis that will not be reported in the survey. To gauge the magnitude of this type of error, we recalculate the incidence of false negatives requiring at least two separate diagnoses of a disease in the OHIP records to be counted as a positive. The results (reported in Table 3d) reveal an average reduction in the incidence of false negatives of 22 percent. The larger decreases are for arthritis (42 percent) and migraines (32 percent) while the smaller are for cataracts (7 percent) and ulcers (8 percent). It is not clear, however, whether this exercise is only compensating for diagnostic errors in the OHIP data. As we show below, the accuracy of respondents’ self reports is positively correlated with the severity of their condition, as measured as the number of OHIP records recording their disease. This result, which has been reported in other studies, suggests we may also be compensating for errors of recall in the NPHS data. Of course, in requiring two instances of diagnoses in the OHIP data we are coding some individuals who have a particular chronic condition as not having the condition. As a result, the percentage of false positives (columns 3 and 4) increases in this exercise.

As a final check we attempt to determine whether the absolute magnitudes of the errors in table 3a are reasonable. To do this, we 1) document inconsistencies in individuals' responses across questions which effectively ask the same thing, and 2) examine a question in the survey for which the OHIP data are almost surely the 'truth'. In the first exercise we examine the relationship between reports of the use of drugs to treat a given condition and the reports of the condition itself. One of the NPHS questions asks if "in the past month did …take any (of a list) of the following medications?". The answers reveal that 7 percent of the respondents who claimed to be taking insulin claimed not to have diabetes, and 10 percent who claimed to have taken "pills to control diabetes" did not report having diabetes. Sixteen percent who claimed to have taken blood pressure medication did not report hypertension and 19 percent of those claiming to have taken asthma medication did not report having asthma. These errors may be slightly overstated, as there are a few conditions, namely pregnancy related gestational diabetes or hypertension, which would require medication but not be associated with the chronic condition. That said, even within the NPHS there are significant discrepancies in individuals' responses suggesting a considerable amount of error in the report of chronic conditions.

In the second exercise we calculate the error in responses to the question "Have you seen a medical doctor or been an overnight patient in a hospital in the past 12 months".[9] Since the doctor or hospital would need to bill OHIP to be paid, and the reference period for the question is within the bounds of our OHIP data, there is a much smaller chance of "spurious" false negative or false positive responses to this question. Here we find that 37 percent of those who had no OHIP record of any kind reported having been to the doctor or hospital (false positives) and 8 percent of those with OHIP records claimed they had no visits (false negatives). Similarly, 28 percent of respondents who had no record of seeing an eye doctor in the past 12 months claim to have been to seen one, and 11 percent of those who have an OHIP record from an eye doctor claim to have not gone.[10]

The one source of OHIP error we are unable to investigate is the fourth account of false positives: the possibility that the individual was diagnosed before the two-year window of the OHIP data, but currently suffers the long-term consequences of the disease. This sort of error is potentially more a problem for ailments such as arthritis than for ailments such as cancer. As a consequence, the information on false positives for conditions in the former group must be interpreted with care.[11]

Quantifying the Measurement Error in Self-Reported Health

We next estimate the magnitude of the measurement error in self-reported chronic conditions drawing on the framework presented in Bound et al (1994) and Bound et al (2000).

Suppose that we are interested in estimating the model

(1) $$y = X^{*}\beta + \varepsilon$$

---

[9] When computing the false positives and negatives we exclude all non-physician OHIP records (namely health practioners) from the sample. This is necessary to properly match the question in the NPHS which asks about consultations with a medical doctor.

[10] The (generally) higher incidence of false positives here than in table 3c may be related to the fact that this question asks individuals to place an event within a fixed interval in the past, while the questions about chronic ailments ask about current experience.

[11] Note also, that a comparison of tables 3a and 3b reveals that the incidence of false positives falls substantially for ailments such as arthritis using the broader mapping of OHIP codes. Changing the mapping presumably does not account for diagnoses that occurred outside the 2 year window of the OHIP data. Therefore some non-trivial part of the false positives for these ailments almost surely does not arise from this source.

One of the explanatory variables, however, say x_i*, is unobservable. Instead we observe x_i which measures x_i* with error. More specifically

(2) $$x_i = x_i^* + v$$

and v is uncorrelated with ε. If we assume the measurement error is "classical", then v is also uncorrelated with x_i*. It is well known that in this case, and when x_i is the only explanatory variable, the proportional bias in estimating β_i (minus the ratio of the bias to the true β_i) is equal to $var(v)/(var(x_i*)+var(v))$.[12]

In the current case we are interested in the possible measurement error in self-reported chronic conditions, which are dichotomous variables. An individual either reports that they have a particular condition or they don't. In this case any measurement error cannot be classical. If $x_i*=1$ then $x_i - x_i*<=0$. Similarly, *if* $x_i*=0$ then $x_i - x_i*>=0$. Therefore, the errors in binary variables must be mean reverting, i.e., $cov(x_i*,v)<0$. In this case, the proportional bias in estimating β_i is equal to the regression coefficient from a hypothetical regression of v on the set of measured explanatory variables. When x_i is the only explanatory variable, this is just $b_{vx} = cov(v,x_i)/var(x_i)$.[13]

In Table 4 we present estimates of the mean error in the self-reported measures of chronic conditions, as well as estimates of the proportional bias that would result when these variables are used as independent variables in a regression. We present both the noise to total variance ratio as well as estimates of b_{vx}, which we denote b_{vS} where S are the self-reports from the NPHS data.

[12] As noted by Bound et al (2000) the measurement error in x_i will also bias the estimated parameters on the accurately measured explanatory variables.
[13] Note that in the classical case with one explanatory variable $b_{vx} = var(v)/var(x_i*)+var(v)$.

Consistent with the statistics in Table 1, the mean error for many of the chronic conditions is quite substantial, reflecting the fact that the self-reported incidence can be more than double or less than one-half the incidence of the ailment in the OHIP data. For the diseases cancer, diabetes, stroke, cataracts, glaucoma and hypertension, false negatives are more prevalent than false positives so the mean error is negative. For the remaining ailments it is false positives that are more prevalent. Some of the conditions in this latter group, for example migraines and back ailments, are likely more prone to self-diagnosis, which would account for the higher incidence of false positives.

For each of the thirteen conditions the noise to total variance ratio, shown in column five, is quite large. It ranges from 0.348 for diabetes, where there is potentially little room for individuals to lie or be mistaken about their condition, to as high as 0.905 for sinusitis. Most of the estimates lie in the interval 0.4 to 0.6.

Our estimates of b_{vS} are of a similar magnitude, but are both smaller and larger than the corresponding noise to total variance ratios. It is possible to show that when the error is negatively correlated with the true value, then the bias will be less than the "variance ratio" if the variance ratio is less than 0.5. In instances where the variance ratio is greater than 0.5 the bias will be greater than in the classical case. As a consequence, the proportional bias when using these as the sole explanatory variable ranges from as low as 20 percent for diabetes to almost 90 percent for arthritis. Particularly large biases are also estimated for ulcers, back ailments, arthritis, bronchitis, sinusitis and migraines. Smaller, although still sizable, biases are estimated for cancer and glaucoma. Even for a presumably straightforward condition like strokes, the estimated bias is more than 50 percent.

We next investigate the relative contributions of false positives and negatives to the total error variance. We decompose σ_v^2 into three components capturing the contributions of 1) false negatives, 2) false positives and 3) the covariance of the two. In columns 1 and 2 in table 5 we report the proportion of the total variance represented by the first two components. First note that the contribution of the covariance is small in all cases, as the proportions almost add to one. Second, the relative contributions of the false positives and negatives are consistent with the preceding inference. For example, in the case of cancer, false negatives account for 88 percent of the total variance in the error. In the case of arthritis, they account for only 5 percent of the total variance. Other conditions for which we would expect larger false positives, such as migraines, ulcers and back problems, do indeed have relatively small portions of the variance due to false negatives. Conditions such as diabetes, glaucoma and hypertension, where self-diagnosis and cured conditions are less likely, all have larger fractions of the variance due to false negatives.

To determine how the false positives and negatives contribute to the attenuation bias that results from using self-reports of chronic conditions as the (sole) explanatory variable in a regression, we calculate a decomposition of the proportional bias, b_{vS}, due to Aigner (1973). It is equal to

(3) $$\Pr(x^*=1 \mid x=0)+\Pr(x^*=0 \mid x=1),$$

or

(4) $$\frac{\pi_{01}\pi}{\pi_{01}\pi+\pi_{00}(1-\pi)}+\frac{\pi_{10}(1-\pi)}{\pi_{10}(1-\pi)+\pi_{11}\pi}$$

using Bayes' rule. $\pi_{01} = Prob(x=0|x^*=1)$ is the probability of a false negative, $\pi_{10} = Prob(x=1|x^*=0)$ is the probability of a false positive, and $\pi = Prob(x^*=1)$. The first term is a function of false negatives while the second term is a function of false positives. In columns 3-7 of Table 5 we report the breakdown of the bias in this fashion. Perhaps surprisingly, in every case the majority of the bias is due to the term representing the false positives. This might have been expected for conditions such as arthritis or migraines where false positives made the greater contribution to the error variance, but not for ailments such as cancer where false negatives played the greater role. Some intuition for this result can be gained from separately examining the numerators and denominators of equation (4). In the first term the proportion of false negatives is divided by the proportion of the population who do not report the chronic condition. In most instances this denominator is fairly large and close to 1.0. In the second term the numerator is the incidence of false positives in the population while the denominator is the proportion of the population who do report having the chronic condition. While the numerator of this term may be small, so is the denominator, so the effect of the false positives has a large "factor loading".

Given the importance of the false positives to the bias, and the fact that using the broad band mapping of the NPHS responses into the OHIP data reduces the incidence of false positives, it is potentially interesting to see if the bias is smaller using the wider mapping. We have replicated the analysis using this alternative mapping and find that the mean error is now negative for all diseases reflecting the higher incidence of false negatives. Furthermore, the estimates of b_{vS} are actually larger in most cases. This likely reflects the tradeoff when using the wider mapping between decreasing the incidence of false positives and substantially increasing the incidence of false negatives.

A qualification to the inference in this section is that the estimates of the attenuation bias that results from using self-reported chronic conditions, as well as it's decomposition, assume that the primary interest is in the effect of OHIP recorded conditions on the dependent variable.

In many instances our interest is in unobserved work capacity, which presumably both the self-reported variable and the OHIP records measure with error.

The Justification Hypothesis

The literature on self-reported health measures suggests that one reason for mis-reporting is that individuals use health to justify their decision not to work. That is, since health is one of few legitimate reasons to be out of the labor force, individuals who face poor labor market opportunities rationalize their absence from the labor market by reporting poor health. To follow up on this possibility, we examine how self-reported health matches up with the administrative data for workers versus non-workers and for full time versus part-time workers. If individuals with poor labor market opportunities are trying to justify their absence from the labor force we might expect that the incidence of false positives would be higher for non-workers relative to workers, and for part-time workers relative to full time workers.

We estimate linear probability models for each of the chronic conditions of the form

$$(self\ report = 1\ \&\ ohip = 0)_i = \gamma_0 + \gamma_1 work_i + X'\lambda_1 + \eta_{1i} \tag{5}$$

limiting our sample to the potential working age population (16 years of age and older and not currently in school). The dependent variable is equal to 1 if an individual self-reports having a particular chronic condition and the OHIP administrative data has no record of such a condition, and zero otherwise. *Work* is equal to 1 if the individual is currently working and zero otherwise. *X* includes a quartic in age, dummy variables (4) corresponding to educational attainment and dummies for sex, residence in an urban area and marital status. As an alternative specification, we restrict our sample to those who have administrative records indicating that they do not have a chronic condition (OHIP=0). The dependent variable is then whether the individual self-reported having a chronic condition. Finally, we also run a similar regression, restricting the sample to workers and replacing the variable work with a dummy variable equal to one if the individual is working full-time and zero if the individual is working part-time.

The results are reported in Table 6. In the first column we report estimates of equation (5) when we exclude the additional covariates in *X*. The estimated parameter on the dummy variable for work is negative for all 13 chronic conditions and significant for all but migraines. Similar results are obtained when we restrict the sample to individuals with no OHIP record of the relevant chronic condition (column 2). The estimated parameter on the work dummy is negative for all 13 chronic conditions, and once again significant for all but migraines[14].

In columns three and four we add the control variables, *X*, to the regression equation. While the explanatory power of work is reduced, it remains a significant negative predictor of false positive reporting for six of the chronic conditions: migraines, asthma, bronchitis, sinusitis, arthritis and ulcers. These are six conditions for which there is likely a higher potential for personal subjective assessment. We should also note, however, that these are also conditions that there is a higher likelihood that individuals could currently suffer the consequences of a ailment that was diagnosed before the 2 year window of the OHIP data. This error in the OHIP data could be the source of a portion of the false positives for these ailments.

The direct effects of the additional covariates on the incidence of false positives are quite mixed. Age, sex, and level of education are correlated with false reports for several of the conditions, but the signs of the coefficients are not consistent across chronic conditions.

The magnitudes of the coefficients are quite significant when considered as a percentage of all false positives. From column 3 we see that working reduces the probability of falsely

[14] We also test specifications that include all other chronic conditions (as recorded in the OHIP data) as additional covariates. The results remain qualitatively unchanged.

reporting migraines by 2.1 percentage points. While alone this number may seem small, taken as a percentage of all false positives for migraines it represents 28 percent. Similar calculations show that, as percentage of all false positives for the ailment, working decreases the probability of false positives for asthma by 83 percent, bronchitis by 96 percent, sinusitis by 45 percent, arthritis by 29 percent and ulcers by 48 percent.

We next restrict the sample to workers and use a full time dummy as an explanatory variable. The results, reported in columns 5 and 6, are mixed. For the full-time/part-time equivalent of equation (5) the coefficient on working full-time is still negative for 7 of 13 chronic conditions, but the coefficients are insignificant for all but one of the chronic conditions. When we restrict the sample to those individuals who have no record of the particular chronic condition in the OHIP administrative data, the results are similar[15].

The Effects of Intensity

We next investigate how the error in self-reported health varies with the intensity of the chronic condition. As a measure of intensity we use the number of times the individual has an administrative record of being treated for a particular condition. We use three intensity levels: 1-5 records, 6-10 records (*REC6-10*), and more than 10 records (*REC11+*). We again use linear probability models in which a 0/1 indicator of false negatives is the dependent variable and the other covariates are as in equation 5:

(6) $$(self\ report = 0 | ohip = 1)_i = \delta_0 + \delta_1 (REC6-10)_i + \delta_2 (REC11+)_i + X'\lambda_2 + \eta_{2i}$$

[15] In addition, we have rerun the above analysis using the wider mapping of ICD-9 codes and have found no significant change in the results. A table of these regressions is available from the authors upon request.

We restrict our sample to those individuals who have at least one OHIP administrative record for a particular chronic condition. It is not possible to run this regression for sinusitis because no one in the sample had more than 6 administrative records with a diagnosis for this condition.

The results are presented in Table 7. For 10 of the 12 chronic conditions having between 6 and 10 administrative records of a chronic condition is strongly associated with lower incidence of false negatives. The exceptions are cataracts and ulcers, for which the coefficients are not statistically significant. Having more than 10 administrative records is less consistently associated with more accurate reporting. However, it should be noted that the sample of people with more than 10 counts of any one condition is extremely small. In sum, our findings suggest that the greater the intensity of the condition, the more likely self-reported health corresponds with the administrative data.

A Labor Supply Example

Self or medical record reports of chronic conditions are ultimately of limited, intrinsic interest to economists. Rather, it is the effect of the underlying, and likely unobserved, health status on some economic outcome that is more important. A prominent example in the literature is the possible effects of health, as manifest in work capacity, on retirement, or more generally, labor supply behavior. To provide additional context for our results, we examine the relationship between labor market participation and our various measures of health using estimation strategies that are common to studies in this area.

As clearly explained in Bound (1991), the measures of health available in surveys of labor market activity have many limitations. First, presumably all are subject to some degree of measurement error because they are only proxies for underlying work capacity. Second, measures of global health, such as the one available in the NPHS and examined in figures 1a-1m, are further limited by measurement error arising from lack of comparability across respondents. Third, direct measures of work capacity, such as questions on physical limitations on activities at home and work, may themselves be a function of labor market opportunities and thus jointly determined.

The measurement error presumably leads to underestimates (attenuation bias) of the effect of health on labor market activity, while the endogeneity bias presumably leads to overestimates. Furthermore, the resulting biases will "spillover" to any other explanatory variable in the market activity equation that is itself correlated with health.

One response to these problems is to attempt to instrument error prone and possibly endogenous measures of global health or physical limitations with more objective measures such as self-reports of chronic conditions or information on the subsequent mortality of respondents (e.g., Stern 1989). While this approach may provide unbiased estimates of the effects of health on labor supply, as argued by Bound (1991) it will lead to underestimates of the effect of labor market opportunities (e.g., the wage), if these are in turn determinants of the more subjective measures of health.

To document these issues in the NPHS, we use a simple model of labor market participation

$$lfp = \beta_0 + \beta_1 \ln w + \beta_2 h + X'\lambda_3 + \eta_3 \tag{7}$$

where lfp is a 0/1 indicator of current labor market participation, w is a measure of labor market opportunities (e.g., the wage), h is a measure of health and X are other determinants of labor market status which include a quartic in age, marital status, sex, and dummies for residence in an urban area and visible minority status. Unfortunately, the NPHS provides information on "family income" rather than on individual level earnings or wages. Even if current wages were available, however, we would still need to impute the market opportunities of individuals who are not in the labor force. To construct a wage for both participants and non-participants we use data from the 1994 (cross section) public use files of the Survey of Labor and Income Dynamics (SLID). This is a nationally representative survey of the Canadian population that is part of a larger panel data set. The longitudinal data are not released publicly, but annual cross-section samples, with certain variables suppressed, are. We estimate a model of individual hourly wages as a function of education (5 levels), a quartic in age, sex, marital status, and an indicator for visible minorities, using the sample of individuals 16 years and older living in the province of Ontario. The wage rates connected with particular jobs held throughout the year are suppressed, so the wage regression uses a composite annual wage rate aggregated by hours of work on each job. The wage equation is then used to impute wages to individuals in the NPHS sample. We exclude education from X in equation (7), so one way of thinking of the imputed wage is as an index of the market value of the education each individual possesses. Certainly this is a fairly coarse method of imputing labor market opportunities to individuals, which potentially limits the amount we can learn about the parameter on this variable[16].

We consider four different measures of health available in the NPHS. The first is the question on global health discussed in figures 1a-1m. We enter this variable as a 0/1 indicator of poor health. The second is a 0/1 indicator of physical limitations constructed from the question: "Because of a long term physical or mental condition or a health problem are/is…limited in the

[16] We have assessed the sensitivity of our results to our method of imputing wages. First, we tried a "selection corrected" wage equation, in which we identified the Mill's ratio off functional form in the absence of any plausible instruments. Second, we tried imputing wages to non-participants (those who reported that their primary activity was something other than 'working for pay or profit') using a wage equation estimated from the sample of workers in the SLID who worked less than 20 (and 30) weeks per year. In all cases, the coefficients of interest in table 8 are affected only in magnitude--the patterns and relative changes across specifications are very similar.

kind or amount of activity you/he/she can do at home?". The third is an index of the self-reported chronic conditions constructed as the count of the number of ailments reported. Similar indices have been used in past studies (e.g., Dwyer and Mitchell, 1999). This index presumably varies with the severity of these conditions through co-morbidity (i.e. the conditions are not independent). Finally, the fourth is a complementary index constructed from the OHIP records.

In the first panel of table 8 we report OLS estimates of equation (7) using these different measures of health status. There are a number of patterns that are consistent with the findings of previous studies. First the role of labor market opportunities is smaller in specifications that include a measure of health than in specifications that don't (i.e., column 1), although the differences are very small. Second, the point estimates suggest that the effect of wage is greater using more objective measures of health. That is, the wage plays a smaller role when using measures of global health or physical limitations than when using the counts of chronic conditions, although again the differences are unlikely statistically significant given the standard errors. Third, in each case the estimated parameter on the health variable is negative and significant. Finally, the estimated parameters on the self and OHIP reported indices of chronic conditions are very similar.

In the second panel we present a first set of IV estimates, using the index of self-reported chronic conditions to instrument for global health and for the incidence of physical limitations[17]. This strategy of using a more objective measure of health as an instrument for subjective measures of health has been used in the past by Stern (1989) and Bound (1991). If the objective measure is uncorrelated with labor market opportunities and any measurement error in the objective and subjective measures is uncorrelated, this strategy potentially addresses the biases discussed above. Of course the preceding evidence casts some doubt on these assumptions, as the erroneous report of some of these ailments is correlated with labor market status. That said, consistent with previous studies, these results suggest that health plays a much larger role and the wage a smaller role in labor market participation.

In the third panel we present another set of IV estimates, using the index of OHIP reported chronic conditions as an instrument We see the same pattern as before: the effect of wage falls and the effect of health increases relative to the OLS results. However compared to the results in the second panel, both these changes are more pronounced.

In table 9 we repeat the analysis restricting the sample to individuals 40 years of age or older. Here we find that the effect of the wage on labor market participation is much smaller and the effect of health marginally larger than was found in table 8. That said, the same pattern of differences between the OLS and IV estimates is observed. [18]

Overall, the results in table 8 and 9 bear strong similarities to the results reported in previous studies. Again we conclude that the inference from the NPHS should inform researchers using self reported subjective and objective measures of health in other surveys.

Conclusions

While this is not the first study to examine the relationship between objective self-reports of medical conditions and health records, previous studies are subject to two major limitations. First, they involve relatively unrepresentative samples. Second, typically only some subset of the measurement error in the data is investigated (Harlow and Linet, 1989). It is in precisely these two ways that our analysis builds upon the previous literature. Our inference is based on

[17] First stage t-statistics are presented in Appendix C.
[18] We have also run IV regressions using 2 other sets of instruments: dummy variables recording self-reports of the chronic conditions as instruments and dummy variables recording OHIP records of the conditions as instruments. The results are largely the same as in table 8 and table 9.

matched survey and administrative data that are representative of the general population. Using a variety of strategies, we investigate the full range of error in these data.

Our results suggest four conclusions. First, there appears to be considerable error, both false positives and false negatives, in self-reported chronic conditions. Estimates of the error variance to total variance often as large as 0.5 and in some cases as large as 0.8. While there is certainly some error in the OHIP health records that we use here as the "truth", we argue through a series of exercises that it is not driving the majority of the error that we document.

Second, using linear probability models to compare workers and non-workers we find some evidence that the error in the self-report of chronic conditions is related to labor market status. That is, individuals may be using their health status as justification for not working. The most striking example is that of bronchitis where we find working decreases the probability of false positives by 96 percent.

Third, we have found evidence that the probability of false reporting decreases with the intensity of the condition. This reflects, perhaps, the fact that increased communication with the physician leads individuals to greater knowledge of their conditions.

Finally, using a simple model of labor market participation we document the relationship between work and our measures of health. Our results confirm the findings of previous studies. For example, using objective measures of health to instrument for subjective measures leads to larger estimates of the effect of health and smaller estimates of the effect of wages on labor market participation.

Appendix A: Mappings of the Chronic Conditions in the NHPS into ICD-9 Codes.

We use two procedures to map the chronic conditions into the ICD-9 codes. The first simply maps chronic conditions into one of 17 broad groupings of ICD-9 codes as follows:

1) neoplasms-- cancer	140-239
2) endocrine, nutritional and metabolic diseases, and immunity disorders-- diabetes	240-279
3) diseases of the nervous system and sense organs—migraines and strokes, cataracts, and glaucoma	320-389
4) diseases of the circulatory system—hypertension	390-459
5) diseases of the respiratory system—asthma, sinusitis and bronchitis	460-519
6) diseases of the digestive system-- ulcers	520-579
7) diseases of the musculoskeletal system and connective tissue—arthritis, back problems	710-739

The second, more narrow, mapping uses more specific ICD-9 codes, whose description exactly matches the chronic conditions. The mapping is as follows:

1) cancer	140-208, 230-234
2) diabetes	250, 253
3) migraines	346
4) stroke	436, 431, 434
5) asthma	493
6) bronchitis/emphysema	506, 490, 491, 492, 518
7) sinusitis	473
8) arthritis	711, 714, 716
9) back problems	724, 738, 722, 730, 731, 732, 733
10) ulcers	530-534
11) cataracts	366
12) glaucoma	365
13) hypertension	401, 405, 416

Appendix B: Percentage of Reports in Medical Records Matched to Report in Survey Instrument, Various Studies

	NPHS	Krueger (1957)	NCHS (1965)	Madow (1973)	Tretli et al. (1982)
Cancer	0.31	0.59	0.33	0.61	
Diabetes	0.63	0.95	0.62	0.81	0.66
Migraines	0.52			0.62	
Stroke	0.47				0.65
Asthma	0.53	0.99	0.71	0.69	
Bronchitis	0.32		0.65	0.79	
Sinus Problems	0.44	0.72	0.48	1.00	
Arthritis	0.66	1.00	0.48		
Back problems	0.45				
Ulcers	0.41			0.60	
Cataracts	0.43				
Glaucoma	0.37				
Hypertension	0.53	0.63	0.46	0.81	

Percentage of Reports in Survey Instrument Matched to Report in Medical Records, Various Studies

	NPHS	Krueger (1957)	NCHS (1965)	Madow (1973)	Tretli et al. (1982)
Cancer	0.77	0.82	0.80	0.53	
Diabetes	0.82	0.93	0.70	0.98	0.66
Migraines	0.20			0.47	
Stroke	0.50				
Asthma	0.49	0.64		0.49	
Bronchitis	0.13		0.16	0.31	
Sinus Problems	0.05	0.54	0.19	0.21	
Arthritis	0.08	1.00			
Back problems	0.30				
Ulcers	0.19			0.60	
Cataracts	0.52				
Glaucoma	0.74				
Hypertension	0.69	0.72	0.46	0.65	

Appendix C: First Stage t-Statistics for the Labor Market Participation Regressions

	Global Health (0/1)	Physical Limitations	Index of Self-Reported Ailments
IV: Index of Self-Reported Chronic Ailments as Instruments			
Full Sample (Age 16+)	17.85	28.29	
Sub-Sample (Age 40+)	11.47	21.30	
IV: Index of OHIP Reported Chronic Ailments as Instruments			
Full Sample (Age 16+)	10.87	12.93	25.32
Sub-Sample (Age 40+)	8.39	10.50	19.06

Notes: Full sample (age 16+) corresponds to the regressions reported in table 8. Sub-sample (age 40+) corresponds to the regressions reported in table 9.

References

Aigner, D.J., 1973, “Regression With a Binary Independent Variable Subject to Errors of Observation”, *Journal of Econometrics*, 1, 49-59.

Bound, J., 1991, “Self-Reported versus Objective Measures of Health in Retirement Models, *Journal of Human Resources*, 26, 106-138.

Bound, J., Brown, C., Duncan, G.J., and W.L. Rodgers, 1994, “Evidence on the Validity of Cross-Sectional and Longitudinal Labor Market Data”, *Journal of Labor Economics*, 12, 345-368.

Bound, J. and R.V. Burkhauser, 1999, “Economic Analysis of Transfer Programs Targeted on People with Disabilities” in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics* v. 3c, New York: Elsevier, 3417-3528.

Bound, J., Brown, C. and N. Mathiowetz, 2000, “Measurement Error in Survey Data”, *Handbook of Econometrics*, forthcoming.

Currie, J. and B.C. Madrian, 1999, “Health, Health Insurance and the Labor Market” in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics* v. 3c, New York: Elsevier, 3309-3416.

Dwyer, D. and O.S. Mitchell, 1999, “Health Problems as Determinants of Retirement: Are Self-Rated Measures Endogenous?”, *Journal of Health Economics*, 18, 173-193.

Harlow, S.D. and M.S. Linet, 1989, “Agreement Between Questionnaire Data and Medical Records’, *American Journal of Epidemiology*, 129, 233-248.

Hux, J., Ivis, F., Flintoft, V, and A. Bica, 2001, “Diabetes in Ontario: Determination of Prevalence and Incidence Using a Validated Administrative Data Algorithm,” Institute for Clinical and Evaluative Sciences mimeo.

Krueger, D.E., 1957, “Measurement of Prevalence of Chronic Disease by Household Interviews and Clinical Evaluations”, *American Journal of Public Health*, 47, 953-960.

Madow, W.G., 1973, *Net Differences in Interview Data on Chronic Conditions and Information Derived from Medical Records,* Rockville MD: National Center for Health Statistics (Vital and Health Statistics, Series 2, No. 57).

National Center for Health Statistics, 1965, *Health Interview Responses Compared with Medical Records*, Rockville MD: National Center for Health Statistics (Vital and Health Statistics, Series 2, No. 7).

Stern, S., 1989, “Measuring the Effect of Disability on Labor Force Participation”, *Journal of Human Resources*, 24, 361-395

Tretli, S., Lund-Larsen, P.G., and O.P. Foss, 1982, “Reliability of Questionnaire Information on Cardiovascular Disease and Diabetes: Cardiovascular Disease Study in Finnmark County”, *Journal of Epidemiology and Community Health*, 36, 269-273.

Williams, J., Iron K., and K. Wu, 1998, “Estimating the Impact of Arthritis on the Burden of Disability and the Cost of Health Care Services” in *Arthritis and Related Conditions: An ICES Practice Atlas*, Toronto: ICES, 11-26.

Table 1: Summary Statistics for Key Variables

Variable	Mean
Age	43.90
	(19.2)
Males	0.45
	(0.50)
Married	0.53
	(0.50)
White	0.91
	(0.28)
Urban	0.88
	(0.33)
Education	
Less that High School	0.31
	(0.46)
High School	0.15
	(0.36)
Some Post Secondary	0.24
	(0.43)
Diploma	0.16
	(0.37)
University Degree	0.14
	(0.34)
Poor health	0.03
	(0.17)
Limited in home activity	0.17
	(0.37)
Chronic condition:	
Cancer	0.02
	(0.15)
Diabetes	0.03
	(0.18)
Migraines	0.09
	(0.29)
Stroke	0.01
	(0.10)
Asthma	0.07
	(0.25)
Bronchitis	0.04
	(0.20)
Sinusitis	0.05
	(0.22)
Back Problems	0.17
	(0.38)
Arthritis	0.18
	(0.39)
Ulcers	0.04
	(0.19)
Cataracts	0.04
	(0.19)
Glaucoma	0.01
	(0.11)
Hypertension	0.11
	(0.31)

Notes: Source is the 1994 NPHS. Standard deviations in parentheses.

Table 2: Correlations between OHIP responses and Self-Reported Health Responses

Condition	Correlation Between OHIP Records and Self-Reported Conditions
Cancer	0.469
Diabetes	0.706
Migraines	0.286
Stroke	0.479
Asthma	0.475
Bronchitis	0.186
Sinus Problems	0.128
Arthritis	0.165
Back problems	0.231
Ulcers	0.261
Cataracts	0.448
Glaucoma	0.516
Hypertension	0.545

Notes: Source is the 1994 NPHS and linked OHIP data.

Table 3a: Summary of False Negative and False Positive Reporting by Chronic Condition Using Narrow Diagnosis Code Band

OHIP Diagnosis	False Negatives as a Fraction of Positives	False Positives as a Fraction of Negatives	False Positives as a Fraction of Survey Positives
Cancer	69.5	0.5	22.6
Diabetes	37.1	0.7	18.4
Migraines	48.1	7.4	79.9
Stroke	53.1	0.5	50.0
Asthma	47.1	3.7	51.0
Bronchitis	67.5	3.6	86.8
Sinus Problems	56.0	5.1	95.5
Arthritis	34.3	17.4	92.1
Back problems	55.0	14.5	76.6
Ulcers	58.8	3.1	81.0
Cataracts	56.9	2.0	48.3
Glaucoma	62.9	0.3	25.9
Hypertension	47.5	3.9	31.3

Notes: Source is the 1994 NPHS and linked OHIP data. Reported statistics are percentages.

Table 3b: Summary of False Negative and False Positive Reporting by Chronic Condition Using Wide Diagnosis Code Band

OHIP Diagnosis	False Negatives as a Fraction of Positives	False Positives as a Fraction of Negatives	False Positives as a Fraction of Survey Positives
Neoplasm with Cancer	86.3	0.5	19.6
Immune System with Diabetes	82.3	0.6	13.9
Nervous System with Migraines	88.2	6.8	43.7
Circulatory System with Stroke	96.5	0.1	10.9
Respiratory with Asthma	88.9	2.4	17.4
Respiratory with Bronchitis	93.4	1.5	18.4
Respiratory with Sinus Problems	92.5	3.0	27.5
Muscle with Arthritis	63.5	10.6	40.2
Muscle with Back problems	70.6	11.9	48.4
Digestive System with Ulcers	89.8	2.4	51.2
Nervous System with Cataracts	92.7	1.4	20.0
Nervous System with Glaucoma	97.4	0.2	10.3
Circulatory system with Hypertension	65.4	2.8	19.1

Notes: Source is the 1994 NPHS and linked OHIP data. Reported statistics are percentages.

Table 3c: Summary of False Negative and False Positive Reporting by Chronic Condition Using Narrow Diagnosis Code Band and 1 year of administrative data

OHIP Diagnosis	False Negatives as a Fraction of Positives	False Positives as a Fraction of Negatives	False Positives as a Fraction of Survey Positives
Cancer	61.9	0.6	26.5
Diabetes	30.8	0.9	25.6
Migraines	46.2	8.0	87.9
Stroke	41.9	0.6	60.9
Asthma	42.4	4.7	65.2
Bronchitis	64.7	3.8	90.5
Sinus Problems	60.0	5.2	97.5
Arthritis	30.4	17.8	95.4
Back problems	48.3	15.4	84.9
Ulcers	58.8	3.4	87.9
Cataracts	54.9	2.5	61.7
Glaucoma	54.7	0.4	32.8
Hypertension	40.6	4.8	39.4

Notes: Source is the 1994 NPHS and linked OHIP data. Reported statistics are percentages.

Table 3d: Summary of False Negative Reporting by Chronic Condition Using 2 or more diagnoses as a positive

OHIP Diagnosis	False Negatives as a Fraction of Positives	False Positives as a Fraction of Negatives	False Positives as a Fraction of Survey Positives
Cancer	58.2	0.7	30.4
Diabetes	27.6	1.1	30.4
Migraines	32.8	8.1	90.1
Stroke	41.7	0.6	54.4
Asthma	33.1	5.0	71.2
Bronchitis	54.6	3.8	92.1
Sinus Problems	50.0	5.2	98.8
Arthritis	20.0	17.8	95.8
Back problems	44.3	15.7	88.3
Ulcers	54.3	3.5	90.8
Cataracts	52.9	2.8	68.9
Glaucoma	48.7	0.4	34.5
Hypertension	34.7	4.7	39.6

Notes: Source is the 1994 NPHS and linked OHIP data. Reported statistics are percentages.

Table 4: Summary Statistics of the Measurement Error in the NPHS Self Reports of Chronic Conditions

Condition	Mean Self Report (*S*)	Mean Diagnosis (*D*)	Mean Error (*v*)	$\frac{\sigma_v^2}{\sigma_D^2+\sigma_v^2}$	b_{vS}
Cancer	0.022	0.056	-0.034	0.447	0.265
	(0.147)	(0.230)	(0.207)		(0.020)
Diabetes	0.034	0.044	-0.010	0.348	0.201
	(0.182)	(0.206)	(0.150)		(0.012)
Migraine	0.089	0.034	0.055	0.717	0.817
	(0.285)	(0.183)	(0.291)		(0.009)
Stroke	0.010	0.011	-0.001	0.503	0.506
	(0.099)	(0.102)	(0.103)		(0.013)
Asthma	0.068	0.063	0.005	0.521	0.542
	(0.252)	(0.244)	(0.254)		(0.013)
Bronchitis	0.041	0.017	0.024	0.739	0.880
	(0.199)	(0.128)	(0.215)		(0.009)
Sinusitis	0.053	0.005	0.047	0.905	0.958
	(0.224)	(0.073)	(0.226)		(0.005)
Arthritis	0.184	0.022	0.162	0.875	0.931
	(0.388)	(0.147)	(0.389)		(0.005)
Back	0.172	0.089	0.082	0.681	0.825
	(0.377)	(0.285)	(0.417)		(0.011)
Ulcers	0.038	0.017	0.020	0.703	0.821
	(0.190)	(0.130)	(0.201)		(0.010)
Cataract	0.039	0.047	-0.008	0.505	0.511
	(0.194)	(0.211)	(0.213)		(0.143)
Glaucoma	0.013	0.025	-0.013	0.436	0.275
	(0.111)	(0.156)	(0.137)		(0.018)
Hypertension	0.108	0.141	-0.033	0.451	0.388
	(0.310)	(0.348)	(0.316)		(0.014)

Notes: Source is the 1994 NPHS and linked OHIP data. Standard errors in parentheses. 3. "Self Reports" are the self-reports from the NPHS while "Diagnoses" are the OHIP medical records. b_{vS} is defined in the text.

Table 5: Decompositions of the Measurement Error in the NPHS Data

Condition	Decomposition of the Error Variance		Decomposition of the Proportional Bias				
	Percent Due to False Positives	Percent Due to False Negatives	Bias: b_{vS}	Due to False Positives	Due to False Negatives	Percent Due to False Positives	Percent Due to False Negatives
Cancer	0.116	0.875	0.265 (0.020)	0.225	0.040	0.850	0.150
Diabetes	0.276	0.715	0.201 (0.012)	0.184	0.017	0.915	0.085
Migraine	0.780	0.193	0.817 (0.009)	0.799	0.018	0.978	0.022
Stroke	0.467	0.527	0.506 (0.013)	0.500	0.006	0.989	0.011
Asthma	0.520	0.448	0.542 (0.013)	0.509	0.032	0.941	0.059
Bronchitis	0.742	0.240	0.880 (0.009)	0.868	0.012	0.987	0.013
Sinusitis	0.935	0.059	0.958 (0.005)	0.955	0.003	0.997	0.003
Arthritis	0.933	0.050	0.931 (0.005)	0.921	0.009	0.990	0.010
Back	0.584	0.331	0.825 (0.011)	0.695	0.077	0.900	0.100
Ulcers	0.735	0.250	0.821 (0.010)	0.810	0.011	0.987	0.013
Cataract	0.407	0.571	0.511 (0.143)	0.483	0.028	0.949	0.054
Glaucoma	0.171	0.823	0.275 (0.018)	0.259	0.016	0.942	0.058
Hypertension	0.328	0.627	0.388 (0.014)	0.313	0.075	0.807	0.193

Notes: Source is the 1994 NPHS and linked OHIP data. Standard errors in parentheses. b_{vS} is defined in the text. Decompositions of error variance do not add to 1.0 because the covariance term is not reported. The decomposition of the proportional bias is due to Aigner (1973).

Table 6: Justification Hypothesis Regressions

Dependent Variable	Self Report=1 & OHIP=0	Self Report=1 | OHIP=0	Self Report=1 & OHIP=0	Self Report=1 | OHIP=0	Self report=1 & OHIP=0	Self Report=1 | OHIP=0
Covariates:	N	N	Y	Y	Y	Y
Coefficient on:	Work	Work	Work	Work	Fulltime	Fulltime
Cancer	-0.005*	-0.006**	-0.001	-0.001	0.004	0.004
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Diabetes	-0.008**	-0.009**	-0.005	-0.005	-0.002	-0.002
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Migraine	-0.006	-0.007	-0.021*	-0.025**	-0.005	-0.005
	(0.008)	(0.009)	(0.011)	(0.012)	(0.017)	(0.018)
Stroke	-0.011**	-0.011**	-0.004	-0.004	-0.004	-0.004
	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)
Asthma	-0.013**	-0.014**	-0.031**	-0.033**	0.004	0.005
	(0.006)	(0.006)	(0.008)	(0.009)	(0.009)	(0.010)
Bronchitis	-0.034**	-0.036**	-0.031**	-0.032**	-0.014	-0.014
	(0.006)	(0.006)	(0.008)	(0.009)	(0.011)	(0.011)
Sinusitis	-0.027**	-0.027**	-0.018*	-0.018*	0.020*	0.020*
	(0.007)	(0.007)	(0.009)	(0.010)	(0.011)	(0.011)
Arthritis	-0.237**	-0.249**	-0.050**	-0.052**	-0.011	-0.009
	(0.012)	(0.013)	(0.014)	(0.015)	(0.017)	(0.017)
Back	-0.037**	-0.051**	-0.018	-0.026*	0.011	0.011
	(0.011)	(0.012)	(0.013)	(0.015)	(0.017)	(0.019)
Ulcers	-0.019**	-0.020**	-0.015**	-0.015**	0.004	0.004
	(0.006)	(0.006)	(0.007)	(0.007)	(0.009)	(0.009)
Cataract	-0.041**	-0.046**	-0.000	0.001	-0.002	-0.002
	(0.005)	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)
Glaucoma	-0.005**	-0.006**	0.002	0.002	-0.001	-0.001
	(0.002)	(0.002)	(0.001)	(0.001)	(0.003)	(0.003)
Hypertension	-0.035**	-0.053**	-0.004	-0.005	0.003	0.005
	(0.006)	(0.008)	(0.008)	(0.009)	(0.009)	(0.010)

Notes: Source is the 1994 NPHS and linked OHIP data. Robust standard errors in parentheses. ** denotes significance at the 5 percent level, * denotes significance at the 10 percent level. The reported statistics are the estimated parameters on a dummy variable for work or a dummy variable for working full time. Where indicated, covariates include a quartic in age, dummy variables for educational attainment (4), sex, residence in an urban area and marital status. Regressions are estimated separately by condition. Sample is restricted to working individuals for the working full time regressions.

Table 7: Estimates of the Effect of Intensity on the Incidence of False Negatives

Condition	6-10 OHIP Diagnoses	11+ OHIP Diagnoses
Cancer	-.408**	-.588**
	(.097)	(.074)
Diabetes	-.301**	-.386**
	(.088)	(.070)
Migraine	-.296**	-.140
	(.125)	(.238)
Stroke	-.428**	-.372
	(.212)	(.236)
Asthma	-.488**	-.082
	(.071)	(.185)
Bronchitis	-.519**	-.255
	(.226`)	(.309)
Sinusitis	-	-
Arthritis	-.420**	-.375**
	(.095)	(.094)
Back	-.251**	-.507**
	(.119)	(.106)
Ulcers	.005	-
	(.181)	
Cataract	-.095	.046
	(.107)	(.200)
Glaucoma	-.333**	-.007
	(.160)	(.318)
Hypertension	-.352**	-.409**
	(.045)	(.049)

Notes: Source is the1994 NPHS and linked OHIP data. Robust standard errors in parentheses. ** denotes significance at the 5 percent level, * denotes significance at the 10 percent level. The reported statistics are the estimated parameters on dummy variables for 6-10 or 11+ reports of the indicated chronic condition in the OHIP data, from a regression of a 0/1 indicator of false negatives on these variables plus a quartic in age, dummy variables for educational attainment (4), sex, residence in an urban area and marital status. Regressions are estimated separately by condition. The estimation sample is individuals who have at least one OHIP administrative record for a particular chronic condition. There were no cases in the data where a respondent had greater than 5 OHIP diagnoses for sinusitis or more than 10 OHIP diagnoses for ulcers.

Table 8: Estimates of the Effects of Wages and Health on Labor Market Participation.

Measure of Health	None	Global Health (0/1)	Physical Limitations	Index of Self-Reported Ailments	Index of OHIP Reported Ailments
OLS					
Wage	0.466	0.447	0.435	0.457	0.461
	(0.037)	(0.037)	(0.037)	(0.037)	(0.037)
Health		-0.312	-0.195	-0.041	-0.033
		(0.032)	(0.017)	(0.008)	(0.009)
IV: Index of Self-Reported Chronic Ailments as Instrument					
Wage		0.403	0.414		
		(0.039)	(0.037)		
Health		-1.018	-0.323		
		(0.149)	(0.038)		
IV: Index of OHIP Reported Chronic Ailments as Instrument					
Wage		0.396	0.393		
		(0.043)	(0.041)		
Health		-1.134	-0.457		
		(0.217)	(0.084)		

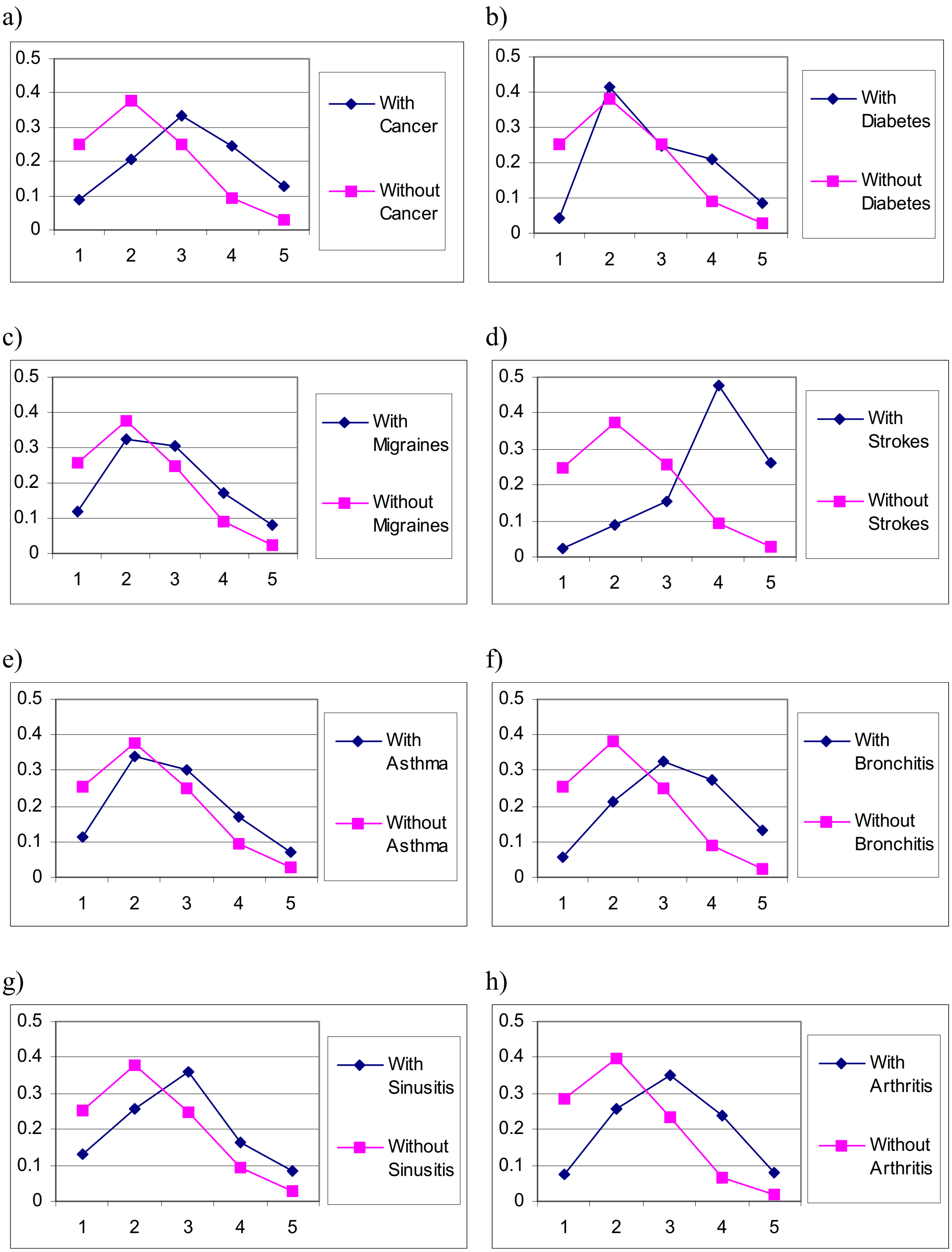
Notes: Source is the 1994 NPHS and linked OHIP data. Robust standard errors in parentheses. The dependent variable is a 0/1 indicator of labor market participation. Other explanatory variables are a quartic in age, marital status, sex, and dummy variables for residence in an urban area and visible minority status. The equation is estimated using the sample of individuals aged 16 and older who are not enrolled in school.

Table 9: Estimates of the Effects of Wages and Health on the Labor Market Participation of Older Individuals.

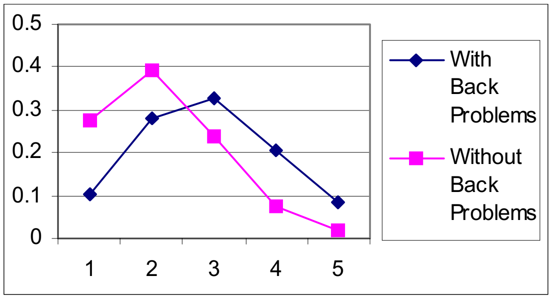
Measure of Health	None	Global Health (0/1)	Physical Limitations	Index of Self-Reported Ailments	Index of OHIP Reported Ailments
OLS					
Wage	0.050	0.047	0.037	0.045	0.046
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Health		-0.313	-0.205	-0.051	-0.042
		(0.036)	(0.020)	(0.010)	(0.010)
IV: Index of Self-Reported Chronic Ailments as Instruments					
Wage		0.040	0.027		
		(0.020)	(0.016)		
Health		-1.232	-0.348		
		(0.217)	(0.045)		
IV: Index of OHIP Reported Chronic Ailments as Instruments					
Wage		0.040	0.018		
		(0.024)	(0.023)		
Health		-1.258	-0.497		
		(0.255)	(0.093)		

Notes: Source is the 1994 NPHS and linked OHIP data. Robust standard errors in parentheses. The dependent variable is a 0/1 indicator of labor market participation. Other explanatory variables are a quartic in age, marital status, sex, and dummy variables for residence in an urban area and visible minority status. The equation is estimated using the sample of individuals aged 40 and older who are not enrolled in school.

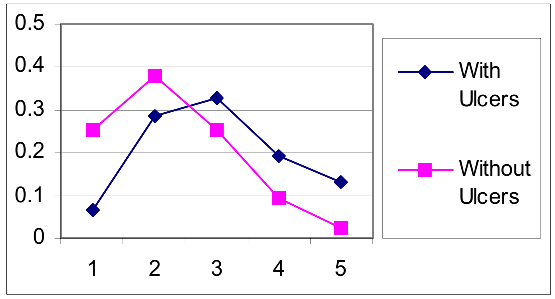
Figures 1: The Distribution of Self-Assessed Health Status by Self-Reported Chronic Condition



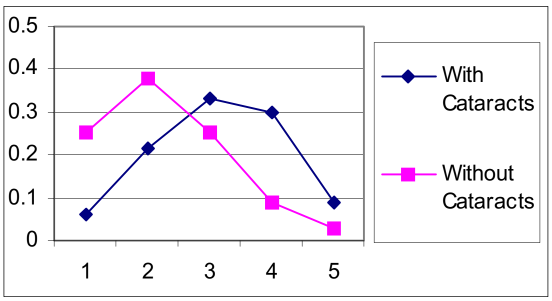
i)



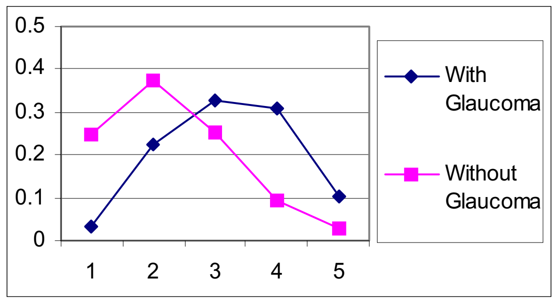
j)



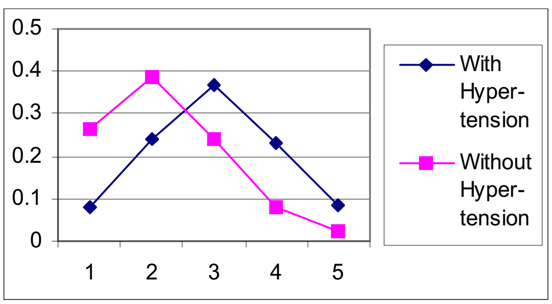
k)



l)



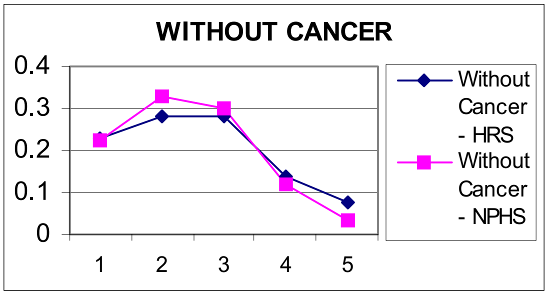
m)



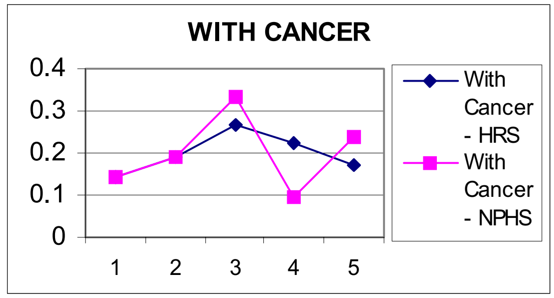
Notes: Source is the 1994 NPHS. The x-axis records self-assessed health ranging from excellent (1) to poor (5). The y-axis records the percent of the sample in each category.

Figures 2: A Comparison of the Distribution of Self-Assessed Health Status by Self-Reported Chronic Condition between the NPHS and HRS.

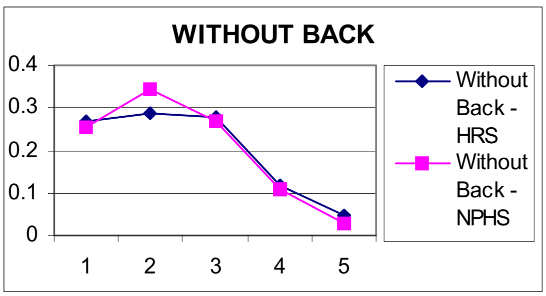
a)



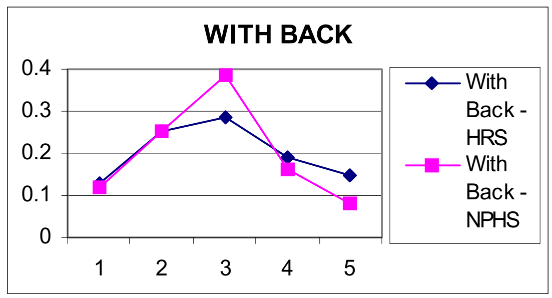
b)



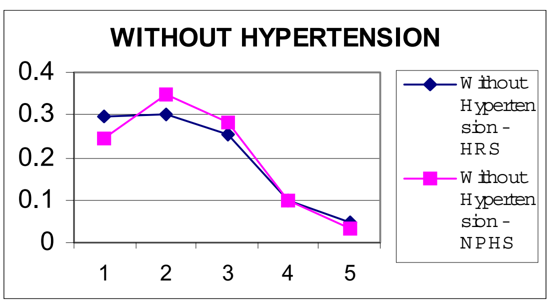
c)



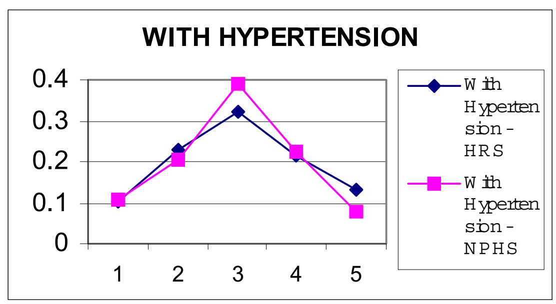
d)



e)



f)



Notes: Source is the 1994 NPHS and 1992 HRS. The x-axis records self-assessed health ranging from excellent (1) to poor (5). The y-axis records the percent of sample in each category.