[4] 厚生労働省大臣官房統計情報部

[5] このように市町村が納付組織を育成していたことを示すものとしては、例えば寿都町国民年金保険料納入組織補助金交付規則（昭和 40 年規則第 2 号）がある。同規則第一条（目的）には、「この規則は、国民年金保険料を自主的に納入しようとする国民年金保険料納入組織の設立奨励及び健全なる事業の発展等を図るために交付する補助金に関し必要な事項を定める」旨規定されている。なお、寿都町は北海道にある水産業中心の町で、その人口は平成 17 年国勢調査によると 3,744 人であり、同規則が制定された昭和 40 年（約 8,000 人）当時から右肩下がりに減少している。

[6] 国立社会保障・人口問題研究所

[7] 「平成 19 年度版　やさしい国税通則法」（2007、川田剛、財団法人大蔵財務協会）

[8] 「平成 19 年度版　やさしい国税徴収法」（2007、山岡千秋、財団法人大蔵財務協会）

参考文献

和泉徹彦（2007a）「年金知識とその形成」『年金制度と個人のオーナーシップ』NIRA 研究報告書 0605, 総合研究開発機構.

和泉徹彦（2007b）「年金通知システムと資産形成」『年金制度と個人のオーナーシップ』NIRA 研究報告書 0605, 総合研究開発機構.

臼杵政治・中嶋邦夫・北村智紀(2008)「厚生年金制度における通知の送付とその効果」『季刊社会保障研究』第 44 巻第 2 号

上村敏之(2007)「年金不信と公的年金の世代間不公平」『年金制度と個人のオーナーシップ』NIRA 研究報告書 0605, 総合研究開発機構.

岡田章（1996）『ゲーム理論』有斐閣.

小椋正立・角田保（2000）「世帯データによる社会保険料負担の納付と徴収に関する分析」『経済研究』Vol.51,No.2.

金子宏 (2007)『租税法　第 13 版』弘文堂.

京極高宣（2007）『社会保障と日本経済』慶応義塾大学出版会.

駒村康平・丸山桂（2005）「国民年金の空洞化問題と年金制度のあり方」城戸喜子・駒村康平編著『社会保障の新たな制度設計』慶應義塾大学出版会.

酒井正（2008）「就業移動と社会保険の非加入行動の関係」国立社会保障・人口問題研究所ディスカッション・ペーパー, No.2008-J01.

菅桂太（2007）「年金未加入と生活保護モラルハザードに関する実証分析」『年金制度と個人のオーナーシップ』NIRA 研究報告書 0605, 総合研究開発機構.

鈴木光男（1994）『ゲーム理論入門』共立出版.

鈴木亘（2008）「医療と生活保護」阿部彩・國枝繁樹・鈴木亘・林正義著『生活保護の経済分析』東京大学出版会.

瀧澤弘和（2007）「第 1 号被保険者による国民年金基礎年金拠出行動のゲーム理論的分析」『年金制度と個人のオーナーシップ』NIRA 研究報告書 0605, 総合研究開発機構.

前田正道（2004）『ワークブック法制執務　全訂』ぎょうせい.

増井良啓（2002）「税務執行の理論」『ファイナンシャル・レビュー』第 65 号.

渡辺智之（2003）「税務回避の経済学：不完備契約としての租税法」『ファイナンシャル・レビュー』第 69 号.

渡辺智之（2005）「租税回避をめぐる問題点」『租税研究』第 672 号.

Sandmo(2005)"The Theory of Tax Evasion: A Retrospective View", *National Tax Journal*, Vol.58, No.4.

Stiglitz(1985)"The General Theory of Tax Avoidance", *National Tax Journal*, Vol.38, No.3.

ROSEN,H.S. and T.Gayer(2007), *PUBLIC FINANCE*, 8th edition, Irwin Professional Publishing Company.

厚生労働科学研究費補助金（政策科学総合研究事業(政策科学推進研究事業)）

「所得・資産・消費と社会保険料・税の関係に着目した

社会保障の給付と負担の在り方に関する研究」

分担研究報告書

「Does Health Status Matter to People's Retirement Decision in Japan? : An Evaluation of "Justification Hypothesis" and Measurement Errors in Subjective Health」

研究協力者　濱秋純哉　東京大学大学院経済学研究科博士課程

研究分担者　野口晴子　国立社会保障・人口問題研究所社会保障基礎理論研究部第２室長

研究要旨

この論文では，健康状態が高齢者の引退行動に与える影響の推定と，得られた健康の効果に含まれうる内生バイアスの大きさの評価を行う。内生バイアスの原因として，(1)回答者が，自分が働いていないことを正当化するために，健康状態を低めに申告する行動(正当化仮説)，(2)健康指標に含まれる測定誤差，の二つに注目する。健康の効果として，就業状態と労働時間への影響を推定するが，これらを推定する際には，操作変数法と操作変数を用いない推定法の両方を行い，両者の結果の違いに基づいて内生バイアスの大きさの評価やその原因の特定を試みる。正当化仮説が成り立っていることによって健康の効果が過大に推定されている場合，操作変数を用いてバイアスを補正すれば，効果は小さくなることが予想される。逆に，測定誤差の存在によって健康指標の係数にゼロ方向のバイアスが生じているなら，操作変数を使うことで健康の効果は大きく推定されるようになるはずである。また，回答者の主観に基づく健康指標と比較的客観的な健康指標の間でどのような推定結果の違いが見られるかによっても，内生バイアスの原因を評価できる可能性がある。推定には2007年度に本研究プロジェクトで行われた第一回「健康と引退に関する調査」の個票データを用いる。推定の結果，健康状態の良い回答者ほど就業している確率が高く，労働時間も長いことが分かった。しかし，内生バイアスの大きさについては，操作変数の妥当性が低いという問題(weak identification)により明確な結論が得られなかった。このため，真の健康の効果がどの程度の大きさなのかについても留保が必要である。したがって，健康の効果を正確に推定するためには，より妥当性の高い操作変数を見つける必要があり，また，より詳細な引退行動に関する情報が必要である。

A. 研究目的

本論文の目的は，健康状態が高齢者の引退行動に与える影響の推定と，得られた健康の効果に含まれうる内生バイアスの大きさの評価を行うことである。内生バイアスの原因として，(1) 回答者が，自分が働いていないことを正当化するために，健康状態を低めに申告する行動 (正当化仮説)，(2) 健康指標に含まれる測定誤差，の二つに注目した。

B. 研究方法

2007 年度「健康と引退に関する調査」の個票データに基づいて，操作変数による推定法 (2SLS, IV probit, IV Tobit) と操作変数を用いない方法 (LPM, probit, tobit) の両方で健康の効果を推定し，両者の結果を比較することで内生バイアスの大きさの評価やその原因の特定を試みた。本論文で用いた健康指標の操作変数は，回答者の居住地が属する二次医療圏の 1 平方メートルあたりの診療所数，回答者の 30 歳時点における運動習慣の有無 (ダミー変数)，回答者の 30 歳時点の BMI (Body Mass Index)，の三つである。

C. 研究結果

操作変数を用いない方法では，健康状態の良い回答者ほど就業している確率が高く，労働時間も長いという結果になった。しかし，操作変数法を用いた推定では，操作変数と健康指標の間の相関が低いために，weak identification の問題が生じ，健康の効果は有意に推定されなかった。

D. 考察

今回用いた操作変数は妥当性が低いために，内生バイアスの大きさやその原因を特定することは難しい。このため，真の健康の効果がどの程度の大きさなのかについては今後の研究で明らかにする必要がある。

E. 結論

健康の効果を正確に推定するためには，より妥当性の高い操作変数を見つける必要がある。

第一回「健康と引退に関する調査」では，引退に関する情報が不足していたが，2008 年度の第二回の調査では引退に関して詳細な質問項目を用意した。それらの情報と二年間のパネル構造を活用することで，より正確に高齢者の健康と引退行動の関係を分析することが可能になると考えられる。

F. 健康危険情報

なし

G. 研究発表

1.論文発表

なし

2.学会発表

・Does Health Status Matter to People's

Retirement Decision in Japan?: An Evaluation of "Justification Hypothesis" and Measurement Errors in Subjective Health, International Health Economics Association, 7th World Congress on Health Economics（於北京，2009年7月12-15日，報告予定）。

H. 知的所有権の取得状況の出願・登録状況

1.特許取得

なし

2.実用新案登録

なし

3.その他

なし

The studies on the design of social security benefit and contribution schemes with attention to the relations between income, assets, consumption and the burdens of social security premium and tax:
Report for Fiscal 2008 (Study Supported by the Health Science Research Grants from the Ministry of Health, Labour and Welfare (Study Project for Promotion of Policy Sciences))

Does health status matter to people's retirement decision in Japan?: An evaluation of "justification hypothesis" and measurement errors in subjective health

Junya Hamaaki
Graduate School of Economics, University of Tokyo
and
Haruko Noguchi
National Institute of Population and Social Security Research

1 Introduction

The Japanese society has been facing a rapid aging and a decrease in a birthrate for the last couple of decades. The large shortage of workforce will be one of the most critical socio-economic issues. Lately, numbers of health economists focus on the relation between health status and job continuation around retirement age, since healthy elderly persons are expected to offset the lack of labor force.

This study has two objectives. First, we evaluate health effects on retirement behavior. Second, we identify the significance of the endogeneity biases in the estimated health effects. We address the biases arose from the following two causes: (1) a tendency of those who are not working to justify their leaving labor force (e.g. early retirement) by false poor health, and (2) measurement errors in the health variables. The phenomenon of (1) is called "justification hypothesis" (Chirikos and Nestel, 1984; Anderson and Burkhauser, 1985; Bazzoli, 1985; Bound, 1991; Waidmann et al., 1995; Dwyer and Mitchell, 1999).

In order to validate the effects of health status on people's choice whether or not they continue to be in labor force after the retirement, we use two empirical approaches: (1) non-instrumented and (2) instrumented regressions. The former includes linear probability (OLS), univariate probit, and Tobit models. Meanwhile, 2SLS, IV probit, and IV Tobit models are in the latter group. Three indicators might be appropriate as IVs as follows: body mass index in 30 years old, the habit of working out in 30 years old, and variations in the number of clinics among different spheres. Further, we use several kinds of health measurements such as binary subjective health, number of diseases, and health status scoring based on principle component analysis. Compared to being not-instrumented, instrumented health effects will be smaller if people actually justify their unemployed status by poor health, and larger if the measurement error is a serious problem in the model. Also, we can evaluate the significance of endogeneity bias by comparing the results between objective and subjective health measures. In this study, we apply a Japanese version of Health and Retirement Survey conducted by the National Institute of Population and Social Security Research in 2007, which was funded by a research grant from the Ministry of Health, Labor and Welfare. The survey focuses on those who are around retirement age and includes detailed information on various objective and subjective health conditions, retirement behavior, job status, working hours, and financial status.

First, this study finds that healthy people are more likely to delay their retirement and work for more hours than those in poor health. Second, we cannot evaluate the severity of endogeneity problems only from the results in this paper because our instruments may be weak. Specifically, 2SLS regression of subjective health measure fails to pass the first stage F-test; moreover, standard errors are much larger in all IV than not-instrumented regressions. Therefore, we conclude that further careful evaluations on the property of self-assessment health measures are inevitable for understanding health effects on people's retirement choice.

This paper is organized as follows. Section 2 discusses the sources of endogeneity bias in self-assessed health status. Section 3 describes our data and its variables, particularly retirement variables and health measures. Section 4 presents the empirical strategy of this paper. Section 5 shows the results of our analyses. Section 6 presents the conclusion of this paper.

2 Sources of endogeneity biases in health status

In the U.S. and Europe, a number of studies have thus far confirmed endogeneity biases of self-reported health status to the outcomes of regressions. One of the reasons for the biases is random measurement error, which gives the coefficient of health status an attenuation bias, i.e. a bias toward zero. Another important source of bias is justification of early retirement by one's false poor health. If this phenomenon is true, poorer health can be observable more likely for both retirees and those who reduced their working hours considerably, all other things being equal. Hence, health effects on labor market participation can be overestimated, as opposed to the case of measurement error.

The bias derived from "justification hypothesis" would be more severe in Japan than other countries where early retirement is not unusual. In fact, according to OECD estimates, the effective retirement age in the period 2002 to 2007, which is defined as a weighted average of net withdrawals from the labor market at different ages over a five year period for workers initially aged 40 and over, is very high both for the male and female workers of Japan (69.5 years and 66.5 years, respectively), compared to OECD average of 63.5 for male and 62.3 for female.[1] Although some part of this gap can be

[1]These OECD estimates are downloadable from the following URL: http://www.oecd.org/dataoecd/3/2/39371902.xls.

explained by the difference in the degree of generosity of public pension system among those countries, the high retirement age indicates that working until old age is widely regarded as a virtue in Japan. Retired people of Japan may therefore be more likely to justify their early retirement by poor health.

Several Japanese works have estimated the relationship between health status and some work-related variables. However, only few works compare effects of objective and subjective health status on wage income, labor market participation, and/or retirement behavior, adjusting the variations in health indices, and controling for the severity of their endogeneity bias (Iwamoto [2000] and Oishi [2000]). Both studies obtain different estimation results between objective and subjective health measurements. In particular, Iwamoto (2000) points out that subjective health indicators (self-rated health and presence of work limitations) have more obvious effects on income and employment, suggesting that the measurement error in subjective health is not so severe that we can use it in empirical analysis. However, this clearer effect of subjective health may be an evidence of the seriousness of the justification bias, which can exaggerate health effects. We therefore try to identify the magnitude of the biases driven by the "justification hypothesis" as well as the random measurement error by utilizing several objective and subjective health indicators.

3 Data and variables

3.1 Data source

The data in this paper is the first wave of the *Survey on Health and Retirement*, conducted by the National Institute of Population and Social Security Research in March of 2007. In order to examine various effects of people's health status on retirement behevior, the survey focuses on males and females who are 45 and older and younger than 80 years old. The 1,074 (525 for males and 549 for females) respondents are randomly extracted out of the 39,311 monitoring samples (19,126 for males and 20,185 for females) owned by the *Central Research Services, Inc* (CRS). The monitoring samples are collected by the monthly omnibus survey conducted by CRS. The CRS extracts samples randomly from the residents' administrative registration records every month and creates the master sample including those who agree to be monitored for all kinds of surveys. For adjusting the distributions of respondents' sex and age to the National Census, the CRS carefully extracts the samples in a way that the number of respondents becomes proportional to the number of population in each sex and 5-year age group based on the residents' administrative registration records in each municipal city. The remuneration paid for respondents is a 500 yen coupon ticket for purchasing books.

3.2 Retirement variables

This paper uses two variables as a proxy of retirement status. One of the retirement variable is a dichotomous variable that takes unity if the individual "has already retired" or "has no jobs." The *Survey on Health and Retirement* asked respondents to choose their employment status from the following eight alternatives: (1) working as a regular employee, (2) working as a contract employee, (3) working as a part-time worker, (4) carrying on business on my own, (5) working at home (e.g. doing side job), (6) other types of work, (7) no job (including housewife), and (8) has already retired after reaching the mandatory retirement age or by early retirement program.[2]

Figure 1 provides the age-specific ratio by sex of those who already retired or have no job, and it shows that the ratio increases with age in both sexes. Although Figure 1 exhibits a similar upward looking, the participation rates in the labor force could be greatly different between males and females. Figures 2 and 3 present the age-specific ratios of "regular employee" and "part-time worker," respectively. Figure 2 tells that the ratio of regular employee for males is considerably higher than females. In addition, Figure 3 shows a higher ratio of part-time worker before 60 years for females than males. These results indicate that women are more likely to choose flexible working styles than men. Moreover, Figure 3 presents a hump-shaped curve after 60 years for males though the ratio for females decreases monotonically, suggesting that a proportion of men in their 60s may be reemployed as a part-time worker after reaching the mandatory retirement age. Thus, the process of retirement differs substantially between males and females; therefore, we can expect different outcomes of the effects of health measures on retirement variable by sex.

Unfortunately, we cannot exactly estimate a health effect on retirement decision in our regressions because health status at retirement is not available in the data; instead, we will examine a health effect on employment status at the survey date. If we consider that individual employment status depends on health conditions at each period, the analysis on the relationship between them is valid as an estimation of health effect on retirement decision. However, the present employment status may be greatly affected by the past retirement experience (including a mandatory retirement), which cannot be controled in our following regression due to the data limitaion. After all, we can only obtain the effect of the current health condition on the employment status at the time.

In order to analyze in more depth the health effect on retirement, we use hours of work per week as an alternative variable that describes the retirement process. This variable is continuous, and it therefore can describe the intermediate retirement status, contrary to the employment dummy variable that considers only the two different ends of employment status between working and retirement. Figure 4 shows average working hours at each age, and it reveals that working hours do not decline drastically even after the general mandatory retirement age of 60 years both for male and female. This decreasing trend of hours worked suggests that people gradually proceed to "full retirement" over their 60s and 70s.

---

[2]Those whose last job is other than regular employee or housewife probably choose (7) because any retirement program is not prepared for them in most cases. Therefore, if the job is regular worker just before withdrawing from labor market, people are expected to choose (8).

3.3 Health measures

This paper uses the follwing three health measures in estimated equations: (1) binary subjective health, (2) the number of diseases, and (3) health status scoring based on principle component analysis.

The binary subjective health made by dividing a five-level subjective health into poor and good health. The five-level subjective health is collected in the following question: "How is your present health status?," which prepares the five alternatives as follows: "Good," "Fairly good," "Neither good nor bad," "Not so good," and "Not good." The binary variable takes unity if the subjective health is 4 or 5, and it also takes zero if the subjective health is 1 or 2 or 3.[3] This conversion helps us to conduct an econometric analysis below though all parts of information are not fully utilized. Another reason for using the binary health status is to avoid a potential bias caused by a small number of observations that choose "Not good."[4]

We use the number of chronic diseases as a more objective indicator of health status than the binary subjective health. This measure counts the number of chronic diseases that the individual respondents were having at the time of the survey. The respondents chose their chronic diseases from a list of 21 diseases (including "other diseases"). This measure does not consider the seriousness of individual medical conditions and their impact on the retirement decision. In other words, all diseases are weighted equally in the calculation of this index. However, this measure can be more likely to be high because a paient with severe disease tend to be affected with other diseases, as suggested in Dwyer and Mitchell (1999). In fact, the people that have severe disease (e.g. heart disease and apoplexy) are rated at a high score of over 3.3, compared to the mean score of 1.5.

Other than subjective health status and number of diseases, we create an alternative health measurement. First, we classified 21 diseases into 5 groups, applying multivariate analyses such as hierarchical cluster analysis and multi dimensional scaling. Hierarchical cluster analysis is a method to measure "dissimilarity" based on Euclidean distances among variables under no external constraints, which is shown in Appendix Figure 1. In Appendix Figure 1, the horizontal axis represents rescaled distance based on standardized squared Euclidean distance (therefore, "dissimilarity") across disease dummy variables. Each disease variable indicates whether or not a respondent has been diagnosed as a specific disease, e.g. heart diseases and joint disorder. Here, we define diseases into the same group, where the rescaled distances are less than 10. In order to verify the results based on hierarchical cluster analysis, multi dimensional scaling is used, which is a way of mapping similar variables to be near and dissimilar ones to be located at a distance. Appendix Figure 2 shows the result. In this figure, the dimensions are also calculated by standardized squared Euclidean distance. When examining individual characteristics

[3] As shown in Appendix Figure 3, the distribution of "health score," whose derivation is explained below, is similar for "Good" or "Fairly good" and "Neither good nor bad." Meanwhile, the distribution of "Not so good" or "Not good" differs from them. Therefore, the inclusion of "Neither good nor bad" in the good health category may be appropriate.

[4] A mere about 3% of observations chose "Not good" in our sample.

in each disease, we find two axes which are (1) relatively high/low levels of body mass index and (2) relatively old/young age of incidence. The results indicate that diseases are similarly classified into 5 groups in both figures. Second, for entire diseases and within each of 5 disease groups, we calculate "disease scores" for each respondent by extracting "principal component" from principal component analysis. The number of diseases shows just how many diseases a respondent was diagnosed so that does not reflect the correlation between diseases. Alternatively, principal component analysis is a convenient statistical method to compose every two-dimensional correlation of multifarious variables into a single synthesized measurement, which takes account of the relations of characteristics of these variables. Besides, while either subjective health status or number of diseases is discrete, principal component "disease scores" are continuous.[5]

The above subjective and objective health measures are highly correlated with each other. Figure 5 shows distributions of the five-level subjective health status by the number of chronic diseases. The distributions indicate that people having more number of diseases tend to be in poorer health than those who have less number of diseases. Moreover, Figure 6 shows the distributions of subjective health status every inter-quartile ranges of principle component disease score. This figure also suggests that the self-assessed health deterioraes as the objective health measure gets worse.

4 The empirical strategy

In what follows, we will describe econometric models and instrumental variables used in IV estimations.

4.1 Empirical formulations

This paper uses two different econometric models to accomplish the following two purposes: (1) estimation of health effects on withdrawal from the labor force, and (2) evaluation of the severity of endogeneity problem in self-reported health status. We will employ the probit model and the linear probability model (LPM) for those two purposes, respectively. We can expect that an attenuation bias will occur in LPM if measurement errors exist in the health status though the direction of the bias is theoretically ambiguous in the probit model.[6] Hence, we can more easily check the seriousness of endogeneity bias

[5]Supplemental explanations of the disease classification and the distribution of the score are given in Appendix.

[6]Levine (1985) considers the measurement error bias in the maximum likelihood estimate (MLE), including probit and censorship type model estimates. He suggests that MLE is affected not only by the classical attenuation bias but also by the additional effects which determine the direction of the bias due to measurement error, differently from normal linear model. Hsiao (1991) and Wang (1998) explore identification conditions for binary choice and censored models, respectively. Two- or three-step procedures for estimating a consistent estimate and the corresponding asymptotic covariance matrices are proposed in their papers. Recently, Edgerton and Jochumzen (2003) reveal by the Monte Carlo and empirical studies that attenuation occurs in the coefficient of independent variable(s) of probit model that is measured with error. They also derive multi-step LIML estimator and find its consistency and

comparing the outcomes of LPM and IV regression than those of the probit and IV probit. However, since LPM has some deficiencies (e.g. some of the LPM fitted value may be outside the unit interval), we use the probit models in order to obtain a precise health effect.

We specify the following four econometric models: (1) univariate probit model, (2) instrumental variable probit model, (3) Tobit model, and (4) instrumental variable Tobit model. Employment status which are dichotomous and working hours which are censored at zero are used as dependent variables in (1)/(2) and (3)/(4), respectively. Here, we omit a specification of LPM because it is the simple OLS that has a dichotomous variable in the left-hand side.[7]

First, we explain the probit models. Let y_i be a binary variable taking unity if the individual have retired or have no jobs and zero otherwise. Consider following binary choice model:

$$y_i = 1(y_i^* = \alpha h_i + \beta X_{1i} + \epsilon_i > 0), \tag{1}$$

where y_i^* denotes an unobserved latent variable; h_i, an observed health measure; X_{1i}, a vector of other household characteristics; and ϵ, a stochastic error term that has a standard normal distribution. If h_i is exogenous precise proxy of unobserved 'true' health, α will be estimated to be consistent. However, if h_i is measured with error, the attenuation bias will occur in the estimate of α. Moreover, under the "justification hypothesis," the health effect on retirement can be overestimated because people try to justify their early retirement by false poor health.

In order to address those endogeneity biases, we employ instrumental variable probit models, including recursive bivariate model. The IV probit model is formulated as follows:

$$\begin{aligned} y_i &= 1(y_i^* = h_i\alpha + X_{1i}\beta + \epsilon_i > 0), & (2)\\ h_i &= X_{1i}\gamma + X_{2i}\delta + v_i, & (3) \end{aligned}$$

where X_{2i} is a vector of additional instruments and (ϵ_i, v_i) has a zero-mean and bivariate normal distribution. The error terms are permitted to be correlated one another, $Cov(\epsilon_i, v_i) = \rho$. On the other hand, this simultaneous model breaks into two parts for y_i and h_i when $\rho = 0$, implying that it is appropriate to use the univariate probit model, eq. (1). Even if h_i is a binary endogenous variable, the above simultaneous model will still generate a consistent estimate, but the result may not efficient. In this case we use the recursive bivariate probit model, wherein the first-stage equation (eq. 2) is a reduced form probit model for binary health indicator, in order to obtain an efficient estimate.

good small-sample property under some assumptions.

[7]Since a binary response in the left-hand side, y, is a Bernoulli random variable in LPM, its conditional variance is expressed as $X\beta(1 - X\beta)$, where X and β are a vector of covariates and its coefficients, respectively. Apparently, heteroskedasticity is present in this variance unless all coefficients are zero; therefore, we use heteroskedasticity-robust standard errors in LPM to deal with this issue.

Next, we show the Tobit models, which are employed to estimate the health effect on hours worked. Let y_i denote working hours, and then we can formulate the standard censored Tobit model as follows:

$$
\begin{aligned}
y_i^* &= \alpha h_i + \beta X_{1i} + \epsilon_i, &(4)\\
y_i &= \max(0, y_i^*), &(5)
\end{aligned}
$$

where y_i^* is a latent variable that is observed for values greater than 0 and censored otherwise; and $\epsilon_i \sim N(0, \sigma^2)$. As in the above probit model, the estimate of α will be biased when the endogeneity occurs in h_i. Suppose we now allow h_i be endogenous. The Tobit model becomes

$$
\begin{aligned}
y_i &= \max(0, \alpha h_i + \beta X_{1i} + \epsilon_i), &(6)\\
h_i &= X_{1i}\gamma + X_{2i}\delta + v_i, &(7)
\end{aligned}
$$

where (ϵ_i, v_i) has a zero-mean and bivariate normal distribution and these terms can be correlated.[8]

4.2 Instrumental variables

In order to address the endogeneity biases of health measures, we use the following three instrumental variables: (1) the density of clinics in a medical sphere where each respondent is located, (2) the experience of fitness habits in one's 30 years, and (3) the deviation of body mass index (BMI) in one's 30 years from its standard value of 22.

The density of clinics in a secondary medical sphere is defined as the number of clinics per square meter in the sphere. We do not use the number of hospitals but use that of clinics because many respondents in our survey chose the individually-managed clinic as their most frequently-used medical institution. The secondary medical sphere is a regional unit of healthcare planning in Japan. Each secondary medical sphere consists of several neighboring municipalities that are socially and geographically related to each other, whereas each primary medical sphere corresponds to a municipality. Since the system of medical service is almost completed within each secondary medical sphere except for the specialty or highly-advanced medical care, the density of clinics at individual district can influence the health of residents.[9]

Second, we describe fitness habits in the past. In the survey, a question about fitness habits, "Did you used to exercise regularly in your 30 years?," is asked. The dummy variable of fitness habits takes unity if respondent answers "Yes" to this question. Since lack of exercise is one of the primary reasons for life-style related diseases (e.g. adiposity,

[8]The first stage regression of IV Tobit for binary subjective health adopts LPM in this paper.

[9]We use the category of the secondary medical sphere at October 1, 2007, wherein there are 355 medical spheres throughout Japan.

diabetes, high-blood pressure, and hyperlipemia), this dummy variable is expected to be significant for health status. In particular, a regular exercise from one's young age is very important for the prevention of those diseases.

Third, the deviation of the past BMI from its standard value can also have a significant effect on the incidence of some diseases. This variable is calculated from a respondent's height in the present (assuming that his or her height has not changed much after 30 years old) and weight in 30 years of age. Respondents whose BMI was higher than 22 in 30 years are expected to be more likely to have life-style related diseases in old age. According to the recent literature of epidemiological research, overweight and obesity in young adulthood and middle age, measured by BMI, are associated with subsequent higher morbidity and disability (Taylor and Østbye, 2001; Ferraro et al., 2002), lower quality of life in older age (Daviglus et al., 2003), higher medicare expenditures (Daviglus et al., 2004; Daviglus, 2005), and later life walking limitation (Stenholm et al., 2007).

The instruments are good predictors of the present health conditional on X_{1i} but to uncorrelate with ϵ_i. The scatter plot figures of instruments confirm their relevance to health conditions. Figures 7-1 to 7-3 plot the mean values of the number of clinics and health measures at each age by sex. They suggest that an increase in the number of clinics in the local area where a respondent lives reduces considerably the number of chronic diseases and the principal component score for males. Moreover, Figures 8-1 to 8-3 give the age-specific ratio of experiencing fitness habits in 30 years and the mean value of health measure at each age, and they indicate that fitness habits in the young decrease the number of chronic diseases and the principal component score for female. Although those negative relations do not appear in males, a U-shaped pattern is found for binary subjective health. Furthermore, Figures 9-1 to 9-3 show the relation between BMI in 30 years and health measures. A high BMI in the past seems to deteriorate the present subjective health whereas an apparent relation is not confirmed for other objective proxies.

Also, those instruments hardly have any correlation with retirement behavior and working hours. Since the density of clinics is an aggregate index of secondary medical spheres, it should not relate with individual working styles. Further, the fitness habits and degree of obesity in 30 years would not correlate with the work behaviors at the middle to old ages.

5 Findings

5.1 Descriptive analysis

In this subsection, we examine the association between health status and work-related variables; i.e. employment status and working hours, using the cross tabulations of Tables 1 and 2. Table 1 provides the variation in the ratio of those who have already retired or have no jobs, working hours, and health status across ten-year age groups for males. In the following analyses, we focus on males because retirement decision and working hours of female workers, particularly when they are married, may be considerably affected by noneconomic factors that is out of our analytical framework. Table 1 does not show a clear correlation between health problem and the work-related variables for men in their 50s probably because the group of poor health includes hard workers who tend to abuse their health as well as retiree and light workers. This result confirms a reverse causality from hard working to the deterioration of health, and thus, this relation asks us to use instrumental variables in our regression analysis. On the other hand, Table 2 shows an obvious negative relationship between health status and working hours for men in their 60s and 70s, and it also shows a positive relationship between health status and retirement ratio. Those results may indicate that a deterioration of health status contributes to a decrease in working hours and the probability of labor force paticipation for the elderly.[10]

5.2 Regression analysis

In this section, we discuss the estimation results of the aforementioned regressions. Before discussing the health effect on retirement, we evaluate the seriousness of endogeneity problem by comparing the results of LPM and 2SLS. If "justification hypothesis" is well applicable to our data, the effect of health status would be overestimated in LPM. On the other hand, if the attenuation bias by measurement errors is relatively important, the health effect expands in 2SLS.

The estimation results of the health effect on binary employment status are reported in Table 3. The effect of binary subjective health is insignificant both in LPM and 2SLS regressions.[11] Meanwhile, the coefficients of two objective health indicators are estimated to be significantly positive in LPM (but not in 2SLS). Of the three 2SLS estimations in column (B), weak identification problem is strongly suspected in the first row because the null hypothesis that the coefficients of the excluded instruments are jointly zero is not significantly rejected. Hence, we focus the outcomes in the second and third rows to evaluate the endogeneity biases. The coefficients of the number of diseases and disease score become insignificant if those measures are instrumented in column (B). This change is consistent with "justification hypothesis"; moreover, it implies that the significance of the estimates in column (A) may be attributed to the endogeneity bias.

However, weak instruments would also be a problem for the objective health measures. Our larger standard errors in column (B) than (A) provide a signal of that problem because 2SLS standard errors for the coefficients of endogenous regressors are much larger than OLS standard errors. One remedy for this problem is to confine the number of instruments if there are more instruments than needed. Therefore, we drop two of three instruments used in Table 3 and re-estimate IV model using only one instrument that seems to be

[10] It must be noted that we cannot completely remove the effect of aging from the variation of numbers in Table 1 even if we divide samples into three age groups. Hence, the relationships between health status and the work-related variables may include a spurious correlation induced by aging.

[11] One of the reasons for this insignificance may be a time gap between subjective health and employment status. The former is a current health status; but, a current unemployed status is a result of one's past retirement decision that may be affected by the health at that time. This mismatch may yield their tenuous linkage.

most strongly correlated with endogenous health variable.[12] However, newly estimated model still yields insignificant coefficients of individual health variables and large standard errors, as presented in Appendix Table 1. Judging from those results, we may not be able to avoid the weak identification problem yet.

Also, we try to identify the possible sources of endogeneity problem based only on the results of LPM because the null hypothesis that the health variable is not endogenous is not rejected for all health measures. Only the coefficients of objective health indicators are significant in column (A). Moreover, the coefficient of subjective health seems to be smaller than it should be because the deterioration of subjective health, which is expressed as a change from zero to unity by binary variable, only has almost the same effect on employment status as a unit increase in the number of diseases and the disease score does. This relatively small effect can be interpreted as a symptom of measurement error in the self-assessed health.

We now evaluate the health effect on retirement based on the univariate probit model in column (C) of Table 3 because we cannot rely on the results of IV probit due to weak identification problem. According to the estimate in the first row of column (C), the binary subjective health does not have a significant effect on the current employment status. On the other hand, the marginal effect of the number of chronic diseases suggests that a unit increase in the number of diseases tend to increase the probability of being retired or having no jobs for males by 4.5 percentage point. Further, if male respondents who have one unit higher "disease scores" are more likely to be retired by 7 percentage points than less diseased males.

Next, we consider the effect of health conditions on working hours. Table 4 presents two different marginal effects of health status calculated in Tobit and IV Tobit models, i.e. the partial derivative of (1) the expected value of the dependent variable conditional on being uncensored, $E(y|y>0)$, and (2) that of the unconditional expected value of the dependent variable, $E(y)$. According to Cameron and Trivedi (2005), those derivatives correspond to the effect of a change in health status on "actual hours of work for workers" and "actual hours of work for workers and nonworkers," respectively.

In columns (A) and (B) of Table 4, poorer status of both subjective and objective health significantly reduces working hours; however, any significant effect is not found in columns (C) and (D). This contrast may also be resulted from the weak identification as pointed out in Table 3. In particular for the binary subjective health, the test of the coefficients of the excluded instruments cannot satisfy the rank condition. Further, the standard errors in IV Tobit model are much larger than that in simple Tobit model, implying that the weak identification problem is extremely severe. Even if we use the strongest instrument variable, IV Tobit regression still yields insignificant marginal effects that are complicated to interpret (see Appendx Table 1). Furthermore, $\rho = 0$ cannot be rejected in Table 4. If we attempt to interpret the results based on the not-instrumented Tobit model, it is difficult to identify whether measurement errors and "justification hypothesis"

[12] We choose the density of clinics as the instrument for the binary subjective health because it is most significant in the first stage regression of Table 3. The past BMI value is used for the number of diseases and the disease score.

prejudice the estimates because health effect on hours worked is statistically significant in all the three subjective and objective measures.

The health effect on hours worked can be calculated as follows from columns (A) and (B). The aggravation of subjective health reduces 4.6 hours of work for workers and 6.4 hours of work for workers and nonworkers. Further, a unit increase in the number of diseases and the "disease scores" decreases 2.4 and 3.6 hours of work for workers, respectively. Also, the increases entail 3.2 and 4.9 hours reductions for workers and nonworkers, respectively.

Table 3 may imply a presence of measurement error in self-assesed health status; but at the same time, Table 4 cannot confirm it. Hence, we should not evaluate the seriousness of endogeneity biases only from our results in this paper due to the complication from the serious weak identification problem. If one seeks to assess this problem further, more valid instruments and more accurate information on retirement behavior will be required.

6 Conclusion

We explore two things in this paper. First, we evaluate health effects on working behavior of the elderly. Second, we try to identify a potential source of endogeneity biases in self-assessed health status. Several significant effects are observed in non-instrumented estimations, and they suggest that poorer health induces early retirement and less working hours. However, we cannot remove endogeneity biases from the health effects because a weak identification problem occurs in our IV regressions. For the same reason, we do not ascertain whether the endogeneity problems are important for our estimates. Therefore, we have to find an appropriate instrument in order to examine the source of endogeneity in health measures more closely.

For a further analysis we add to second wave of *Survey on Health and Retirement* questions about retirement including its experience, reason, date, expected retiment age, and amount of retirement allowance. Moreover, questions about the amount of public pension benefits or its expected value are also asked in the second wave because the pension system substantially affects individual retirement behavior. Additionally, the second wave collects more comprehensive information (e.g. health status, retirement experience, retirement allowance, pension benefits) as to respondents' spouses than the first wave in order to control for spouse's behavior and increase the number of observations that can be used for regression analysis. When the data in the second wave of the survey become available, we will be able to conduct some further analyses utilizing additional queries and panel structure.

Appendix. Supplemental remarks on the principal component disease scores

Twenty-one diseases are classified into Group I including heart diseases, joint disorder, eye disease, high blood pressure, diabetes, and osteoporosis; Group II with chronic lung disease and Parkinson's disease; Group III counting urinate disorder incl. incontinence/prostatomegaly and psychological disorder; Group IV including hyperlipemia, cerebrovascular disorder/stroke, and liver disease; and Group V with other diseases which cannot be categorized into any groups (asthma, gastrointestinal disorder, thighbone fracture, ear disease, skin disease, cancer, and other diseases) (see Appendix Figure 1). Appendix Figure 2 shows that both Group I and Group II are characterized by relatively old age of incidence and high level of BMI, Group III by relatively old age of incidence and low level of BMI, Group IV by relatively young age of incidence and high level of BMI, respectively.

Appendix Figure 3 shows the distribution of principal component disease score for all diseases by three kinds of subjective health status, "Good" or "Fairly Good," "Neither good nor poor," and "Not so good" or "Not good." On the horizontal axis, the smaller value of "disease score" to the left shows better health. Obviously, the distributions of disease score seems to be highly skewed to the left in better subjective health group and to become gentle and symmetric in worse health groups. The distributions of disease score for each disease group provide similar results as the above (see Appendix Figures 4-1 to 4-5). The larger variance of disease score may imply larger measurement errors in worse than better subjective health groups.

References

Anderson, Kathryn, H. and Richard, V. Burkhauser (1985) "The Retirement-health nexus: A new measure of an Old Puzzle" *Journal of Human Resources* Vol. 20, No. 3, pp. 315-330.

Bazzoli, Gloria, J (1985) "The Early Retirement Decision: New Empirical Evidence on the Influence of Health" *Journal of Human Resources* Vol. 20, No. 2, pp. 214-234.

Bound, John (1991) "Self-reported versus Objective Measures of Health in Retirement Models" *Journal of Human Resources* Vol. 26, No. 1, pp. 106-138.

Chirikos, Thomas, N. and Gilbert Nestel (1984) "Economic Determinants and Consequences of Self-reported Work Disability" *Journal of Health Economics* Vol. 3, No. 2, pp. 117-136.

Daviglus, M. L., Kian Liu, Lijing, L. Yan, Pirzada Amber, Daniel, B. Garside, Linda Schiffer, Alan, R. Dyer, Philip Greenland, and Jeremiah Stamler (2003) "Body Mass Index in Middle Age and Health-Related Quality of Life in Older Age The Chicago Heart Association Detection Project in Industry Study" *Archive Internal Medicine* Vol. 163, pp. 2448-2455.

——, Kian Liu, Lijing, L. Yan, Pirzada Amber, Larry Manheim, Willard Manning, Daniel, B. Garside, Renwei Wang, Alan, R. Dyer, Philip Greenland, and Jeremiah Stamler (2004) "Relation of Body Mass Index in Young Adulthood and Middle Age to Medicare Expenditures in Older Age" *Journal of American Medical Association* Vol. 292, No. 22, pp. 2743-2749.

—— (2005) "Health Care Costs in Old Age are Related to Overweight and Obesity Earlier in Life" *Health Affairs* (web exclusives).

Dwyer, Debra, S. and Olivia, S. Mitchell (1999) "Health Problems as Determinants of Retirement: Are Self-rated Measures Endogenous?" *Journal of Health Economics* Vol. 18, No. 2, pp. 173-193.

Edgerton, David, and Peter Jochumzen (2003) "Estimation in Binary Choice Models with Measurement Errors" *Lund University, Department of Economics WP* 2003:4.

Ferraro, K. F., Ya-ping Su, Randall, J. Gretebeck, David, R. Black, and Stephen, F. Badylak (2002) "Body Mass Index and Disability in Adulthood: A 20-Year Panel Study" *American Journal of Public Health* Vol. 92. No. 5, pp. 834-840.

Hsiao, Cheng (1991) "Identification and Estimation of Dichotomous Latent Variables Models Using Panel Data" *Review of Economic Studies* Vol. 58, No. 4, pp. 717-731.

Iwamoto, Yasushi (2000) "Health and Income" (Kenko to shotoku), In National Institute of Population and Social Security Research ed. *Changing Functions of Families and Households: Special Attention to the Safety Net for the Japanese Elderly* (Kazoku setai no henyo to seikatsu hosho kinou). Tokyo: University of Tokyo Press. (in Japanese)

Levine, David (1985) "The Sensitivity of MLE to Measurement Error" *Journal of Econometrics* Vol. 28, No. 2, pp. 223-230.

Oishi, Akiko (2000) "Effects of Health on the Participation of Elderly Workers in the Labour Force" (Koreisha no shuugyo kettei ni okeru kenko yoin no eikyo) *Japanese Journal of Labour Studies* (Nihon rodo kenkyu zasshi) Vol. 42, No. 8, pp. 51-62. (in Japanese)

Stenholm, Sari., Päivi Sainio, Taina Rantanen, Seppo Koskinen, Antti Jula, Markku Heliövaara, and Arpo Aromaa (2007) "High Body Mass Index and Physical Impairments as Predictors of Walking Limitation 22 Years Later in Adult Finns" *Journal of Gerontology*, Vol. 62A, No. 8, pp. 859-865.

Taylor, D. H. and Truls Østbye (2001) "The Effect of Middle- and Old-Age Body Mass Index on Short-Term Mortality in Older People" *Journal of the American Geriatrics Society* Vol. 49, pp. 1319-1326.

Waidmann, Timothy, John Bound, Michael Schoenbaum (1995) "The Illusion of Failure: Trends in the Self-Reported Health of the U.S. Elderly" *NBER Working Papers* No. 5017, National Bureau of Economic Research, Inc.

Yatchew, Adonis, and Zvi Griliches (1985) "Error in Probit Models Review of Economics and Statistics" Vol. 67, No. 1, pp. 134-139.

Wang, Liqun (1998) "Estimation of Censored Linear Errors-in-variables Models" *Journal of Econometrics* Vol. 84, No. 2, pp. 383-400.

Figure 1. Ratio of retiree or no-job person

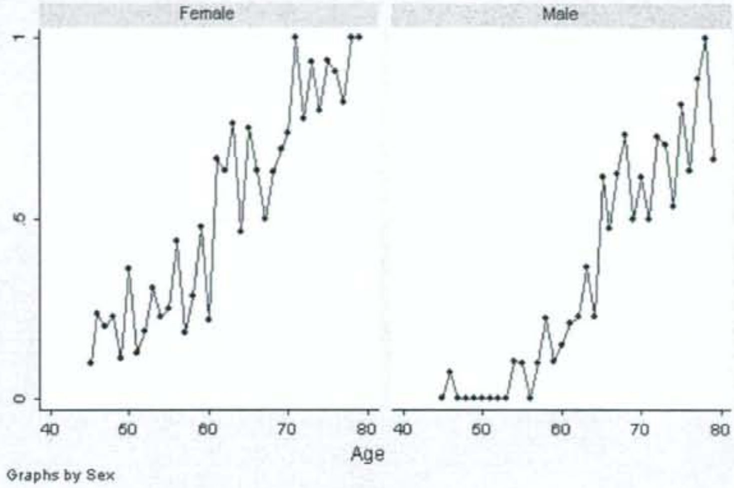

Figure 2. Ratio of regular employee