ment errors in the health variables. First, our objective health measures widely distribute in the poorer subjective health. This indicates a possibility of the endogeneity problems in subjective health. Second, the limitation of daily activities is strongly correlated with one's employment status in spite of its weak correlation with exogenous factors determining health. A likely explanation of this result is that respondents report the poorer health status than true one to justify their employment status or they may not be able to assess their own health status accurately oneself. Third, instrumented health effects are larger than not-instrumented ones, as would be consistent with an alleviation of attenuation bias. Finally, the intrinsically unrelated variables are positively correlated with each other both in time series and cross section. This demonstrates a noticeable tendency of justifying the retirement in good economic condition.

This paper is organized as follows. Section 2 formulates our empirical specification. Section 3 describes the data source and our variables, including employment status variables and health measures. Section 4 looks for the exogenous determinants of health status. Section 5 shows the estimation results. Section 6 adduces an evidence supporting "justification hypothesis." Section 7 presents the conclusions of this paper.

2 Empirical specifications

This paper employs various empirical specifications, including probit, linear probability model (LPM), two-stage least squares (2SLS), and Tobit models, in order to accomplish the evaluation of the endogeneity problems. We can expect that attenuation bias will occur in LPM if measurement errors exist in the subjective health indicator. Meanwhile, the direction of the bias is theoretically ambiguous in the maximum likelihood estimate (MLE).[1] Hence, we compare the outcomes of LPM and 2SLS regressions to check the seriousness of attenuation bias. However, since LPM has some deficiencies (e.g. some of the LPM fitted value may be outside the unit interval), we also use the probit model in appraising "justification hypothesis."

We specify the following three econometric models: (1) univariate probit model, (2) IV probit model, and (3) Tobit model. Dichotomous employment status indicators and censored working hours are dependent variables in (1)-(2) and (3), respectively. Here, we omit a specification of LPM because it is a simple OLS which has a binary variable in the left-hand side.[2]

[1]Levine (1985) considers the measurement error bias in 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 good small-sample property under some assumptions.

[2]Since a binary response in the left-hand side 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,

First, we explain the probit models. Let y_i be a binary employment status variable; for example, it takes unity if an individual has retired or has no jobs and zero otherwise. Consider the following binary choice model:

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

where $y_i^\star$ 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 which has a standard normal distribution. If h_i is an exogenous health variable, 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 IV probit model. This model is formulated as follows:

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

where X_{2i} is a vector of additional instruments and (ϵ_i, υ_i) has a zero-mean and bivariate normal distribution. The error terms are permitted to be correlated one another, $Cov(\epsilon_i, \upsilon_i) = \rho$. 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 estimate may not be efficient. In this case we have to 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.

Next, we show the standard censored Tobit model that is adopted to estimate the health effect on hours worked. Let y_i denote working hours, and then we formulate it as follows:

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

where $y_i^\star$ is a latent variable which is observed for values greater than 0 and censored otherwise; and $\epsilon_i \sim N(0, \sigma^2)$. IV Tobit model allows h_i be endogenous through a correlation of the error terms in the health and working hours equations. However, we do not use it due to a severe weak identification problem in this model.

---

respectively. Apparently, heteroskedasticity is the case we have to consider in this variance unless all coefficients are zero; therefore, we use heteroskedasticity-robust standard errors in LPM to deal with this issue.

3 Data and variables

3.1 Data source and sample selection

The data in this paper is the *Survey on Health and Retirement*, conducted by the National Institute of Population and Social Security Research in March of 2008 and 2009. In order to examine various effects of people's health status on retirement behavior, the survey focuses on males and females who are 45 and older and younger than 80 years old. For the first wave of the survey in the year of 2008, 2,747 people are randomly extracted out of 39,311 monitoring samples 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. Out of those, 1,074 people responded the survey (valid response rates: 39%) in the first wave. Then, the second wave is a follow-up survey on these 1,074 respondents. Out of 1,074, 862 respondents (response rate: 80%) answered the survey and so 212 (approximately 20%) dropped out from the sample. Further, in the second wave, 578 people are newly chosen at random from CRS monitoring samples. Out of 578, 257 people (response rate: 44%) responded the survey.

This survey has a couple of unique characteristics which is different from the data used in previous studies such as Iwamoto (2000) and Oishi (2000). First, the survey asked respondents diseases in detail which were diagnosed by physicians. Hence, we can control for respondents' both subjective and relatively objective chronic health status more accurately than previous works. Second, the survey includes the data on a respondent's retirement and re-employment history in the past. Therefore, this study would distinguish respondents who have not been retired yet from those who have been re-employed either on full-time or part-time basis since the first compulsory retirement.

Among whole sample, we use only male respondents in our econometric analysis because of some complications of the female retirement behavior. Since a number of life-cycle related factors (e.g. getting married, baring children, and providing long-term care to family members) make female workers leave labor market more often compared to male workers, a simple analytic framework probably cannot describe the mechanism of female retirement behavior.[3] Moreover, compared to males, females are less likely to feel embarrassed by leaving labor market in young and having no job, which is not uncommon for women in Japan. Therefore, "justification hypothesis" probably does not matter to female workers. For the same reason, Iwamoto (2000) also does not include female work-

[3] 11.5% of female respondents choose "Other reasons" as a reason for retirement while only 5.3% male respondents choose it. This is an evidence of the variety in the causes of the female retirement.

ers into the empirical work.[4] Consequently, the number of remaining sample is 465 that excludes outliers of health measures.[5]

3.2 Employment status and working hours

This paper uses three kinds of binary variables as a proxy of employment status. The definition of those variables depends on the following ten alternatives in *Survey on Health and Retirement*: (1) a regular employee on full-time basis, (2) a contract worker, (3) a temporary staff (including day worker), (4) a part-time worker, (5) self-employed (including farmer, forestry, and fishery), (6) a freelance profession (e.g. writer), (7) working at home (e.g. doing side business), (8) a skilled worker or profession (e.g. physician or lawyer), (9) other working status, and (10) no job (including a full-time domestic worker or a retired person). The first variable takes unity if the respondent is working as a regular employee ([1]). The second one indicates whether the respondent is working as an irregular employee ([2], [3], and [4]). The last one depends on whether the respondent has already retired or has no jobs ([10]). We compare the possibility of the justification behavior among those three employment status. For example, a comparison between not working as a regular employee and not working at all is an interesting subject of our study. Further, we use hours of work per week as an alternative variable that describes the retirement process. This variable is continuous, and therefore, it can describe the intermediate retirement status, contrary to the employment dummy variable.

In Japan, elderly males often become non-regular employees in a time period between the first compulsory retirement and the time when they left completely from the labor market. Figure 1 shows males' age-specific ratio of employees by type of employment status and hours of work per week. The ratio of regular employees is obviously decreasing, but its slope does not seem to be very steep. The average working hours also do not decline drastically even after the general mandatory retirement age of 60 years. This is because a proportion of the male elderly are likely to be reemployed as a non-regular employee after the mandatory retirement, as described by a hump-shaped curve of the irregular employee. Thus, the retirement ratio gradually approaches to unity as the ratio of regular to non-regular employees decreases, suggesting that people gradually proceed to full retirement over their 60s and 70s. Compared to male workers, the ratio of non-regular out of total employees is much higher than that of regular employees for female workers. Further, the retirement ratio in the earlier 60s for females exceeds 50 percent, compared to 14 percent for male workers. As mentioned in the previous section, females may be unlikely to feel ashamed of her non-regular status and early retirement due to this high retirement ratio.

[4]Oishi (2000) also focuses on the male elderly though the reason is not mentioned clearly.

[5]Specifically, 4 respondents report extreme values in the number of disease (19, 20, and 23). Those values correspond to the largest number of disease score (more than 10). They are excluded from our sample to avoid a bias due to outliers.

3.3 Health measures

This paper uses the following four health measures in estimated equations: (1) poor overall self-rated health, (2) presence of limitations of daily activities at home and/or at work, (3) the number of diseases, and (4) health status scoring based on principle component analysis.

The poor overall self-rated health is a binary health indicator that is constructed by dividing a five-level subjective health into poor and good health. The original multi-level health is collected by the following question: "How is your present health status?," which prepares the five alternatives as follows: (1) "Good," (2) "Fairly good," (3) "Neither good nor bad," (4) "Not so good," and (5) "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). Although some information on health is wasted in this conversion, this simplifies our econometric analysis below. Another reason for using the binary health status is to avoid very few observations in the worst health status.[6]

The second one is a proxy of the presence of limitations of daily activities at home and/or at work. This dummy variable is generated from the following original question: "Did your poor physical and mental conditions hinder your daily life and business in the past year?" Respondents can choose one among (1) "Not at all limited" (2) " Not greatly limited," (3) "Relatively limited," and (4) "Quite limited." The binary variable takes unity if respondents choose (3) or (4), and zero otherwise.

We use the number of chronic diseases from which a respondent has not completely recovered at the time of survey in 2009 as a more objective indicator than the binary subjective health. This measure counts the number of chronic diseases that individual respondents were having at the time of the survey. The respondents chose their chronic diseases from a list of 27 diseases, such as heart disease, hypertension, hyperlipemia, stroke and/or cerebral vascular disorder, cancer and/or malignant tumor, diabetes, gout, chronic lung disease, asthma, hepatitis, renal disorder and/or gallbladder disease, digestive system and/or gastrointestinal disease, fibroid disease and/or ovarian disease, thyroid disease, emiction disorder, joint disorder, hernia, backache, foot fracture, osteoporosis, eye disease, ear disease, pollinosis, parkinsonism, skin disease, depression, digestive system disorder other than gastrointestinal related disease, and other disease. 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 for the patients with severe disease because they tend to be affected with other diseases, as noted in Dwyer and Mitchell (1999, p.179). In fact, the people that suffer from one or more have 2.3 diseases on average.

Other than subjective health status and number of diseases, we create an alternative health measurement. For entire diseases, we calculate "disease scores" for each respondent by extracting "principal component" from principal component analysis. The number of

[6]A mere 1.5% of respondents (7 observations) chose "Not good" in our using sample. In addition, only 5% of respondents selected the worst category in the question of the limitation of daily activities.

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.

The above subjective and objective health measures are highly correlated with each other. Figures 2a and 2b show the distributions of the five-level and four-level subjective health status by the number of chronic diseases. The distributions indicate that people suffering from a number of diseases tend to be in poorer health than those who have less disease. Moreover, Figure 3a and 3b displays the distributions of subjective health status every inter-quartile ranges of principle component disease score. It also suggests that the self-assessment health deteriorates as the objective health measure gets worse.

At the same time, we can point out a clear trend of the disease score distribution according to the self-rated health status. Figures 4a and 4b show the histograms and the fitted kernel density distributions of principal components score. We can observe its wider variations in the poorer health status. If the patients with a bad objective health honestly report their health status, the disease score cannot distribute so widely. Hence, this tendency may be a signal of the difficulty in precisely measuring one's own health or the justification behavior by the falsely-rated poor health. Further, Figures 5a and 5b also ascertain the same trend of the number of chronic diseases. In particular, the number of diseases has two peaks in "quite limited" category of Figure 5b. Also, those figures verify validity of our division of the five-level or four-level rated subjective health into two. A large difference in the shape of their distributions, including the skewness in the poorer categories, is found between the poor and good categories of our indicator.

4 Exogenous determinants of health status

In this section, let us consider the problem: what exogenous variables affect health status? We have two purposes of considering this problem. The first one is to quantify "true" versus "sham" measurements in health status. The second is to search appropriate IVs for health status. Our exogenous variables are the following three variables: (1) distance in a straight line between one's home and the nearest low-volume hospital, (2) the density of clinics in a medical sphere where each respondent is located (per m^2 in a medical sphere), and (3) the body mass index (BMI) in one's 30 years of age. The first two variables are proxies for accessibility to the medical resources. The third one can be a predictor for the health status in one's old age.

In order to calculate the distance IV, we use geographic information system (GIS), specifying the latitudes and longitudes based on zip codes of each respondent's residence and the nearest low-volume hospital and estimating the physical distance in a straight line between them. Here, "low-volume" is defined as a medical facility which has more than

20 beds and less than 100 beds.[7] For respondents, the closer distance to the nearest low-volume hospital implies better accessibility to health care on daily basis, which will lower the opportunity costs to obtain medical care. Therefore, it may have a positive impact on their health status. On the other hand, in particular in rural area, the relation could be reversed such that an elderly person lives closer to medical resources because they are more likely to have trouble in their health. To control for this reverse causality, we add cross terms of the city size dummy variables and the distance. The city size dummies are assigned for 18 major cities (17 ordinance-designated cities and Tokyo's 23 wards) and other cities. The reference group is towns and villages. However, the physical distance to the hospital is less likely to be correlated to people's retirement behavior, only on a relatively strong assumption such that respondents are randomly allocated, regardless of the locations of medical facilities (McClellan and Newhouse, 1997; McClellan and Noguchi, 1998).[8]

The density of clinics in a secondary medical sphere is defined as the number of clinics per square meter in the sphere. We use the number of clinics rather than the one of hospitals because many respondents in our survey chose individually-managed clinics as primary care facility. 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]

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).

Table 2 reports the estimation results of eq. (3). The covariates other than health

[7]In Japan, medical facilities with less than 20 beds are categorized as "clinics," not "hospital."

[8]The elderly might move for better accessibility to medical and long-term care. If many respondents moved because of "bad health," the assumption of randomness on their locations to hospitals would not hold, which means that the distance is not an appropriate IV due to the endogeneity. In order to examine this problem, we have to know whether the health status does matter to respondents' physical address. However, unfortunately, the previous survey did not include the information on mobility. We add the question on the relation of health status with geographical mobility in the follow-up survey which will be conducted in the year of 2010.

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

variables are age, age squared, respondent's income, household wealth, dummy variables for academic background, past retirement experience, and pensioner dummy variable. The results can be summarized as follows: (1) distance to the nearest low-volume hospital significantly worsens elderly persons' health status especially in urban areas, (2) density of clinics in the secondary medical sphere improves objective health status measured by disease score (but its p-value is only 0.124), and (3) past obesity increases the number of chronic diseases and disease score. In relation to (2), the coefficients of the distance in towns and villages are significantly negative. This is consistent with the aforementioned possibility of the reverse causation problem in rural area. On the basis of those results, disease score is most likely to be consistent with the exogenous health-related variables. This may imply that disease score includes more variations of "true" health status than other measures do. Moreover, we can say that the distance to the low-volume hospital and past BMI can be appropriate IVs for the present health status.

5 Findings

This section carries out two empirical analyses to assess the severity of the measurement error problem and "justification hypothesis."

5.1 Comparison of the effects of subjective and objective health

The first analysis compares the properties and effects on employment status and hours worked between subjective and objective health measures. Table 3 reports health effects on three employment status, obtained in the regression of the simple binary choice model (eq. [1]). The explanatory variables are the same as the aforementioned regression of eq. (3). In the rows (A) and (E), subjective poor health seems not to be a useful indicator for determining one's employment status. Second, limitations of daily activities at home and/or at work are strongly correlated with one's employment status, though this measure was quite weakly correlated with the exogenous variables in Table 2. Hence, this measure may reflect the unobserved factors affecting the employment status or "justification hypothesis" holds true as regards this measure. Third, the number of diseases and the disease score also have a significant effect on one's employment status. Fourth, the marginal effects on full-time work are somewhat different from the ones on other employment status with respect to a couple of aspects as follows. First, the effects of subjective poor health have significant impacts only on this status. This result may suggest that people who do not work as a regular employee try to justify their status by underreporting the self-assessment health. Second, in fact, a significant effect is hardly found for the more objective indicators, one of which is most consistent with the "true" health variation in Table 2. This may be a signal of measurement errors even in our objective health measurements or indicates that health status does not matter to regular employment status.

Table 4 reports the results of Tobit model (eqs. [4] and [5]). It presents two different marginal effects of health status, i.e. the partial derivative of (1) the expected value of the dependent variable conditional on being truncated, $E(y, y > 0|x)$, and (2) that of the expected value of the dependent variable conditional on being censored, $E(y|x)$. According to Cameron and Trivedi (2005, p. 542), 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. A striking contrast between the significance of subjective and objective health is one of the most remarkable features in this table. The health effects are not significant for subjective poor health and limitations of daily activities. On the other hand, the number of diseases and disease score have a highly significant effect on the hours worked. A possible reason for this contrast is that the incentives to justify the small number of working hours are not as strong as ones to account for unemployment status. In other words, the justification behavior is clearly observed only in explaining one's leaving labor market. Another possibility is serious measurement errors in the self-assessment health. The wide variation of the number of chronic diseases and disease score in poorer health status, appeared in Figures 4a, 4b, 5a, and 5b, is also consistent with this interpretation. However, this cannot rationally explain the close correlation between limitations of daily activities at home and/or at work and the three employment status in Table 3. Thus, the former possibility may be more applicable to our data than the latter one.

5.2 Comparison of instrumented and not-instrumented estimates of health effects

In this section, we discuss the estimation results of IV regressions. If error terms in the employment and health equations are correlated, the use of instrumental variables can remove any endogeneity biases in the health effect. Judging from the results in Table 2, we can use the distance to the nearest low-volume hospital and BMI value in one's 30 years of age as instruments for disease score. The correlation of other IV than above measures with health status is too weak to pass the diagnostic test of weak identification problem. Although the distance IV closely correlates with the subjective health measures in Table 2, the recursive bivariate probit model generates merely unreliable results.

Table 5 reports IV estimates on the disease score. The size of the coefficients with IVs becomes much larger than those in probit and LPM. This result may indicate an alleviation of the attenuation bias. However, almost all effects are insignificant due to large standard errors, which is an evidence of weak instruments. Since weak identification problem is detected here, we cannot derive a decisive conclusion only based on this result.

6 Evidence of "justification hypothesis" based on seemingly unrelated variables

This section shows an evidence of "justification hypothesis" in a different analytical framework from regressions in the previous section. In fact, we analyze the relation between the following two variables: (1) the retirement due to bad health and (2) a job openings-to-applications ratio (*yuko kyujin bairitsu*). If respondents honestly answer their health status, those two variables should not be related. However, if "justification hypothesis" holds true, those variables can be positively correlated because respondents who retired in the time period of high job openings ratio may be more likely to justify their retirement by false poor health. In calculating the retirement ratio, we use all retired observations (both males and females) that fill in the question; "What is the reason for retiring your last job? (multiple answers allowed)." This question prepares the following thirteen alternatives: (1) getting married, (2) baring children, (3) mandatory retirement, (4) getting laid off or encouragement to retire, (5) applying voluntarily the early retirement program, (6) bankruptcy of the company, (7) receiving pension payments, (8) poor health, (9) providing long-term care to family members, (10) to enjoy spending time with one's family, (11) no need for working, (12) no desire to work, and (13) other reasons.

Figures 6a and 6b show the time-trend of the retirement due to illness and the annual average of the job openings-to-applications ratio (excluding new graduates and including part-time workers).[10] The job openings ratio is taken from the report of the Employment Service Agency (*shokugyou antei gyoumu toukei*). The numbers in Figures 6a and 6b are five-years and three-years average value, respectively.[11] Both figures indicate the time-series positive correlation between those intrinsically unrelated variables, implying the validity of "justification hypothesis." However, this relation appears to be slightly weak in the first several periods in Figure 6b, since respondents would not try to justify the retirement almost a couple of decades ago by bad health. In addition, Appendix figures 1a, 1b, 2a, and 2b exhibit the relations of the retirement due to other two major reasons (mandatory retirement and getting married) and the job openings ratio. Those two variables are not correlated with each other, unlike the retirement due to poor health. This result suggests that the respondents do not justify their retirement by those reasons.[12]

Next, our focus is on the cross-sectional correlation between the abovementioned two variables. Figure 7 plots the ratio of the retirement due to bad health in individual prefectures against the job openings-to-applications ratio. The prefecture-level job openings ratio is the average value during the period from 1980 to 2007. Also, we use only 25 prefectures that contain more than 10 observations of the reason for retirement. As a result, Figure 7 reveals a positive correlation between those two variables. Table 6 reports that the coefficient of the linear slope in that figure is significantly positive at the 10% level. This is our second evidence for "justification hypothesis." Appendix figures 3 and 4 also show the scatter plot of the mandatory retirement and retirement after marriage, respectively. The ratio of the mandatory retirement seems to randomly distribute across

[10]In calculating those average values, we use the data since 1983 because the number of total observations in each year is very small (less than 10) before 1982.

[11]In order to check the robustness of the time trend, we report the two different average values here.

[12]The retirement due to bankruptcy of respondent's company is negatively correlated with the job openings-to-applications ratio. This is a natural relationship because bankruptcy occurs in the economic recession, and it suggests that the respondents honestly report the reason in this case, in contrast to the retirement due to bad health.

the job openings-to-applications ratio, as one would expect. Meanwhile, the retirement after marriage has a close negative correlation with the job openings ratio. This is also confirmed by the significantly negative linear slope in column (C) of Table 6. This downward slope may simply reflect the fact that people can find another job more easily in the prefecture of the higher job openings ratio. Thus, justification behavior is not a common phenomenon of any retirement reasons, but it is unique to the retirement due to illness in our data.

7 Conclusions

Whereas the health effect on the elderly retirement behavior has recently attracted much research attention in Japan, the endogeneity of health measurements would be a problem in estimating it. We tried to identify the significance of the bias arising from "justification hypothesis" and classical measurement errors in this paper. Our strategies are (1) analyzing the properties of various health measurements (distributions, correlation with exogenous variables, and effects on employment status and working hours), (2) using IVs to compare non-instrumented and instrumented health effects, and (3) analyzing the relation of the intrinsically unrelated variables to verify "justification hypothesis."

We find some symptoms of respondents' justification behavior and measurement errors in our health variables. First, the number of diseases and the disease score widely distribute in the poorer subjective health status. This can be a signal for both problems of "justification hypothesis" and the measurement error. If neither problem were not severe, the objective health measures would not distribute so widely. Second, the limitation of daily activities is strongly correlated with one's employment status though it has only weak correlation with exogenous factors determining health. This indicates that respondents justify their leaving labor market by reporting false poor health or that they have a difficulty in measuring their own health, which causes a measurement error problem. Third, the effects of disease score on the employment status expand after instrumented, suggesting that measurement error problem is detected. However, since weak identification problem could potentially occur in our IVs, we cannot derive decisive conclusion from the IV regression. Finally, seemingly unrelated variables have a positive correlation in both time series and cross sectional settings. This is a favorable evidence for the justification behavior in a good business condition.

A limitation of our study is that we could hardly rely on the results of IV regression due to a lack of the appropriate IVs. If we can collect the medical records and the cause of death of the respondents' family members (particularly those of parents) in the third wave of *Survey on Health and Retirement*, they become more suitable IVs than our present ones.

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.

McClellan, M. B., and Haruko Noguchi (1998) "Technological Change in Heart-Disease Treatment: Does High-Tech Mean Low Value? " *American Economic Review* Vol. 88, No. 2, pp. 90-96.

McClellan, M. B. and J. P. Newhouse (1997) "The Marginal Cost-Effectiveness of Medical Technology: A Panel Instrumental-variables Approach," *Journal of Econometrics* Vol. 77, No. 1, pp. 39-64.

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.

Table 1. Top three reasons for past retirement

	Male	Female
1.	Mandatory retirement (53.6%)	Getting married (19.4%)
2.	Bad health (13.5%)	Bad health (14.5%) Mandatory retirement (14.5%)
3.	Bankruptcy of the company (7.2%)	-

Source: Author's calculation from *Survey on Health and Retirement*.

Table 2. Exogenous determinants of health statuses

	Dependent variable	Poor health		Limitation of daily activities		Number of chronic diseases		Disease score	
		Probit				OLS			
		dy/dx		dy/dx		Coef.		Coef.	
(A)	Distance (to the nearest low-volume hospital) (δ_1)	−0.005		0.001		−0.057	*	−0.029	*
		(0.009)		(0.011)		(0.032)		(0.017)	
	Distance*18 major city dummy (δ_2)	0.058	***	0.057	**	0.132		0.075	*
		(0.021)		(0.029)		(0.118)		(0.044)	
	Distance*Other city dummy (δ_3)	0.003		0.006		0.034		0.017	
		(0.011)		(0.013)		(0.040)		(0.019)	
	Test of δ_1=0 & δ_2=0 & δ_3=0								
	p-value	0.0407		0.1066		0.2171		0.1043	
(B)	Density of clinics in a medical sphere	−0.004		0.000		−0.008		−0.008	#
		(0.004)		(0.005)		(0.017)		(0.005)	
(C)	BMI in 30 years old	0.004		−0.008		0.084	*	0.038	**
		(0.007)		(0.009)		(0.044)		(0.016)	

Note: Heteroskedasticity-robust standard errors are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$ # $p<0.15$. Reference group of raw (A) is "towns and villages." Number of observations is 465. Other household characteristics are not reported in this table.

Table 3. Health effects on employment status

	Dependent variable	Full-time work		Part-time work		Not working	
		Probit					
		dy/dx		dy/dx		dy/dx	
(A)	Poor health	-0.130	**	-0.023		0.089	
		(0.046)		(0.032)		(0.092)	
(B)	Limitation of daily activities	-0.140	***	-0.062	**	0.147	*
		(0.044)		(0.025)		(0.088)	
(C)	Number of chronic diseases	-0.031	*	-0.015	*	0.039	*
		(0.017)		(0.008)		(0.039)	
(D)	Disease score	-0.017		-0.051	**	0.093	*
		(0.045)		(0.021)		(0.051)	
		LPM					
		Coef.		Coef.		Coef.	
(E)	Poor health	-0.059	*	-0.019		0.071	
		(0.036)		(0.042)		(0.050)	
(F)	Limitation of daily activities	-0.063	*	-0.079	**	0.090	**
		(0.034)		(0.034)		(0.040)	
(G)	Number of chronic diseases	-0.010		-0.014		0.034	***
		(0.009)		(0.009)		(0.011)	
(H)	Disease score	0.001		-0.055	**	0.087	***
		(0.018)		(0.023)		(0.029)	

Note: Heteroskedasticity-robust standard errors are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Number of observations is 465. Household characteristics are not reported in this table.

Table 4. IV estimates of the health effect on employment status

Dependent variable	Full-time work		Part-time work		Not working	
	IV Probit					
	dy/dx		dy/dx		dy/dx	
Disease score	0.312		-0.209		0.211	
	(0.235)		(0.213)		(0.190)	
First-stage Wald Chi^2	14.16	**	15.59	**	12.85	**
	2SLS					
	Coef.		Coef.		Coef.	
Disease score	0.172		-0.163		0.274	*
	(0.147)		(0.155)		(0.159)	
First-stage F-value	2.21	***	2.21	***	2.21	***
Hansen's J stat (p-value)	0.861		0.9332		0.9666	

Note: Heteroskedasticity-robust standard errors are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Number of observations is 465. Other household characteristics are not reported in this table.

Table 5. Tobit estimates of the health effect on working hours

	Tobit					
	(A) Maginal effect on E(y| x)			(B) Maginal effect on E(y, y>0| x)		
	Coef.	Std. Err.		Coef.	Std. Err.	
Poor health	−2.514	1.969		−3.350	2.659	
Pseudo R^2			0.117			
Limitation of daily activities	−0.995	1.847		−1.316	2.453	
Pseudo R^2			0.117			
Number of chronic diseases	−1.276	0.541	**	−1.678	0.712	**
Pseudo R^2			0.119			
Disease score	−3.596	1.487	**	−4.727	1.959	**
Pseudo R^2			0.119			

Note: Dependent variable is working hours. Std. Err. denotes heteroskedasticity-robust standard error. The marginal effects on $E(y|x)$ and $E(y, y>0|x)$ are calculated as $\partial E(y|x)/\partial x$ and $\partial E(y, y>0|x)/\partial x$ evaluated at mean of other covariates, respectively. For binary health, the marginal effect corresponds to $E(y|x=1)-E(y|x=0)$ and $E(y, y>0|x=1)-E(y, y>0|x=0)$. ** $p<0.05$. Number of observations is 458.

Table 6. The cross-sectional correlation between seemingly unrelated variables

	OLS		Tobit
	(A)	(B)	(C)
Retirement reason	Poor helath	Mandatory retirement	Marriage
	Coef.	Coef.	Coef.
Job openings-to-applications ratio	0.136 *	−0.053	−0.205 **
	(0.070)	(0.070)	(0.099)
Constant	0.035	0.366 ***	0.263 ***
	(0.061)	(0.064)	(0.090)
(Pseudo) R^2	0.116	−0.024	−1.100

Note: Heteroskedasticity-robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Number of observations is 25. R^2s in columns (A) and (B) are adjusted for degrees of freedom.

Figure 1. Ratio of each employment status and hours of work per week by 3-years age groups